Personalizing Many Decisions with High-Dimensional Covariates

Nima Hamidi
Stanford University

Mohsen Bayati
Stanford University

Kapil Gupta
Airbnb

Overview

How to test new medical interventions

1. A hospital wants to reduce hospital acquired infections:
    - E.g., Use one of two newly designed catheters (A or B).
2. They should select one of A or B per patient.
3. A/B test or Randomized Controlled Trial (RCT) have high opportunity cost.
    - In healthcare, experimentation is costly or unethical.[1]

[1] Sibbald, Bonnie. 1998. *Understanding controlled trials: Why are randomized controlled trials important?*, British Medical Journal (Clinical Research Ed.) 316(201).

Beyond healthcare

- Kohavi and Thompke, Harvard Business Review, 2017:

  'Today, Microsoft and several other leading companies – including Amazon, Booking.com, Facebook, and Google – each conduct more than 10,000 online controlled experiments annually, with many tests engaging millions of users.'

- But, experiments have opportunity costs.

Multi-armed bandit experiments

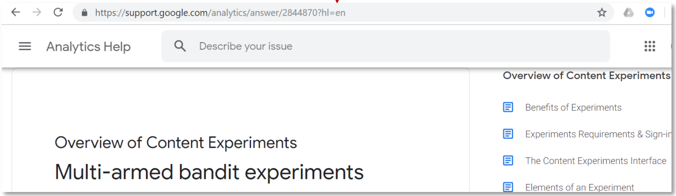
RCT
Pure Exploration

Multi-armed bandits
Balance Exploration & Exploitation

Experimentation is OK

Experimentation is not OK

Thompson'33 and Lai-Robbins'85



Adding Covariates

- Treatment outcomes depend on a set of covariates (context or features).

- E.g., in an A/B testing case, A is optimal for a subset of the patients/users and B is optimal for the remaining ones.

K-armed contextual bandits with linear pay-off

- Patients arrive with covariates $X_t \in \mathbb{R}^d$ where $X_t \sim_{\text{i.i.d.}} \mathcal{P}_X$.
- At time t, reward of arm i is

$$Reward(i; X_t) := X_t^\top B_i + \varepsilon_t.$$

  - B_i's are unknown parameter vectors.
  - ε_t's are sub-Gaussian mean-zero independent.

Formal setting

1. Each arm i corresponds to an **unkown** vector $B_i \in \mathbb{R}^d$.

2. At time t, a **context vector** $X_t \in \mathbb{R}^d$ is revealed to the policy.

3. The policy π selects action $a_t \in [k]$.

4. The **reward** is given by $y_t = \langle B_{a_t}, X_t \rangle + \varepsilon_t$.

Formal setting

1. Each arm i corresponds to an **unkown** vector $B_i \in \mathbb{R}^d$.
2. At time t, a **context vector** $X_t \in \mathbb{R}^d$ is revealed to the policy.
3. The policy π selects action $a_t \in [k]$.
4. The **reward** is given by $y_t = \langle B_{a_t}, X_t \rangle + \varepsilon_t$.

We further assume:

1. X_t's are i.i.d.
2. ε_t's are independent mean-zero sub-Gaussian.
3. $(X_t) \perp\!\!\!\perp (\varepsilon_t)$
4. B is of rank r.

Cumulative regret

Definition

We define the **cumulative regret** of a given policy as follows:

$$R_T = \sum_{t=1}^{T} \left[\max_{1 \le i \le k} \langle B_{t,i}, X_t \rangle - \langle B_{t,a_t}, X_t \rangle \right].$$

Policies with smaller (expected) regrets are desired.

Theoretical guarantees

- OLS-Bandit: $O(d^2 k^3 \log(T))$
- Lasso-Bandit: $O(s^2 k^3 \log(T)^2)$
- REAL-Bandit: $O(r^2 (k+d) \log(T)^2)$

REAL-Bandit

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

REAL-Bandit

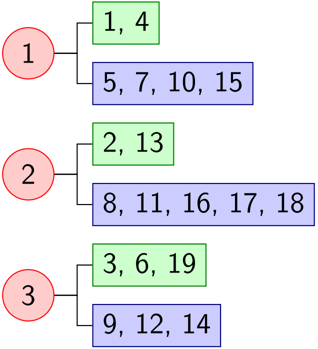
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

REAL-Bandit

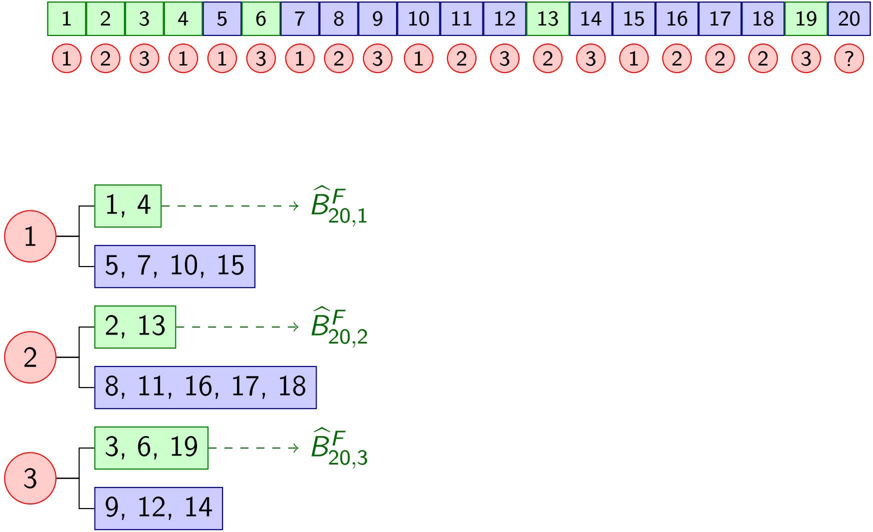
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	1	1	3	1	2	3	1	2	3	2	3	1	2	2	2	3	?

REAL-Bandit

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	1	1	3	1	2	3	1	2	3	2	3	1	2	2	2	3	?



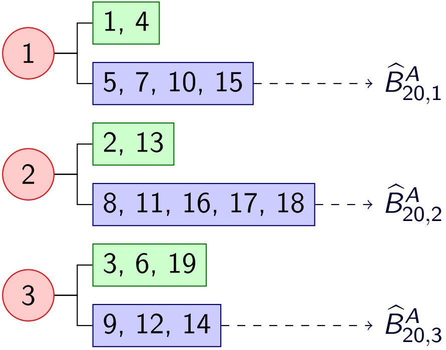
REAL-Bandit



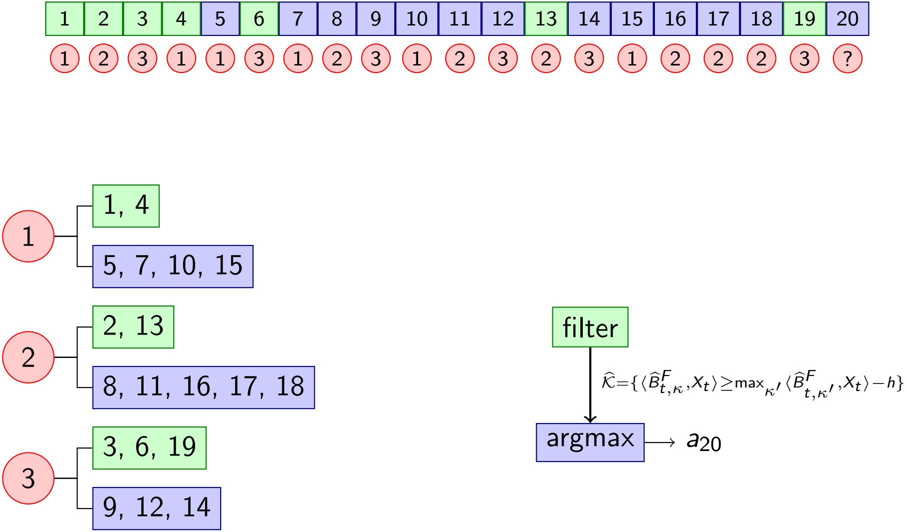
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
1 2 3 1 1 3 1 2 3 1 2 3 2 3 1 2 2 2 3 ?
1
1, 4
$\widehat{B}^F_{20,1}$
5, 7, 10, 15
2
2, 13
$\widehat{B}^F_{20,2}$
8, 11, 16, 17, 18
3
3, 6, 19
$\widehat{B}^F_{20,3}$
9, 12, 14

REAL-Bandit

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	1	1	3	1	2	3	1	2	3	2	3	1	2	2	2	3	?

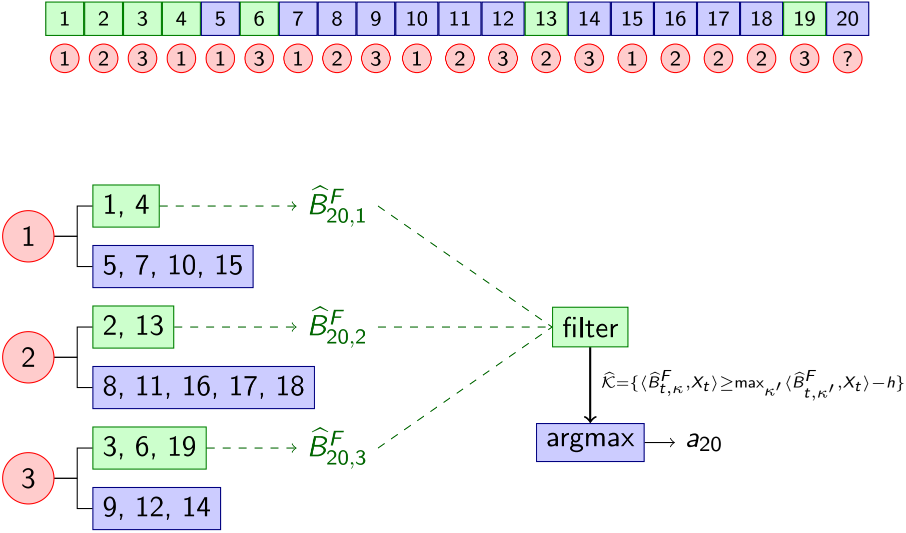


REAL-Bandit

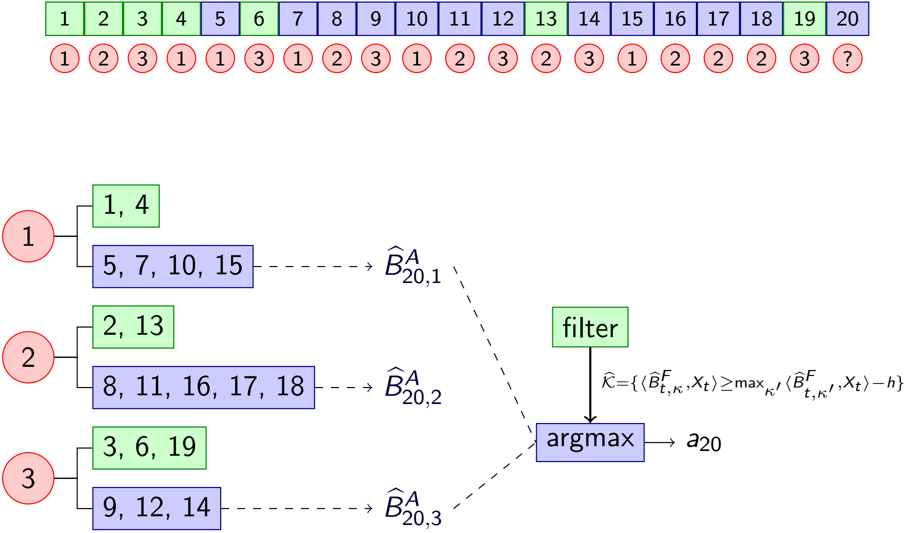


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
1 2 3 1 1 3 1 2 3 1 2 3 2 3 1 2 2 2 3 ?
1
1, 4
5, 7, 10, 15
2
2, 13
8, 11, 16, 17, 18
3
3, 6, 19
9, 12, 14
filter
$\widehat{\mathcal{K}}=\{\langle \widehat{B}^F_{t,\kappa}, X_t\rangle \geq \max_{\kappa'} \langle \widehat{B}^F_{t,\kappa'}, X_t\rangle - h\}$
argmax
a_{20}

REAL-Bandit


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
1 2 3 1 1 3 1 2 3 1 2 3 2 3 1 2 2 2 3 ?
1
1, 4
5, 7, 10, 15
2
2, 13
8, 11, 16, 17, 18
3
3, 6, 19
9, 12, 14
$\widehat{B}^F_{20,1}$
$\widehat{B}^F_{20,2}$
$\widehat{B}^F_{20,3}$
filter
$\widehat{\mathcal{K}}=\{\langle \widehat{B}^F_{t,\kappa}, x_t\rangle \geq \max_{\kappa'} \langle \widehat{B}^F_{t,\kappa'}, x_t\rangle - h\}$
argmax
a_{20}

REAL-Bandit


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
1 2 3 1 1 3 1 2 3 1 2 3 2 3 1 2 2 2 3 ?
1
1, 4
5, 7, 10, 15
$\widehat{B}^{A}_{20,1}$
2
2, 13
8, 11, 16, 17, 18
$\widehat{B}^{A}_{20,2}$
3
3, 6, 19
9, 12, 14
$\widehat{B}^{A}_{20,3}$
filter
$\widehat{\mathcal{K}} = \{\langle \widehat{B}^{F}_{t,\kappa}, x_t\rangle \geq \max_{\kappa'} \langle \widehat{B}^{F}_{t,\kappa'}, x_t\rangle - h\}$
argmax
a_{20}

REAL-Estimator

- Use any low-rank estimator, such as

$$\bar{B} := \arg\min_{B} \frac{\|Y - \mathfrak{X}(B)\|_2^2}{n} + \lambda \|B\|_*$$

such that the following holds with high probability

$$\left\|\bar{B} - B\right\|_F^2 \leq C\sigma^2 \frac{dr}{n}.$$

- This bound leads to extra $\sqrt{k}$ in the regret bound.

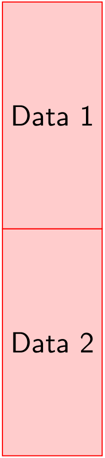
REAL-Estimator

- Let $\bar{B}$ be defined as in the previous slide.
- Run the following "*row-enhancement*" procedure.
- This procedure eliminates extra $\sqrt{k}$ factor in the regret.

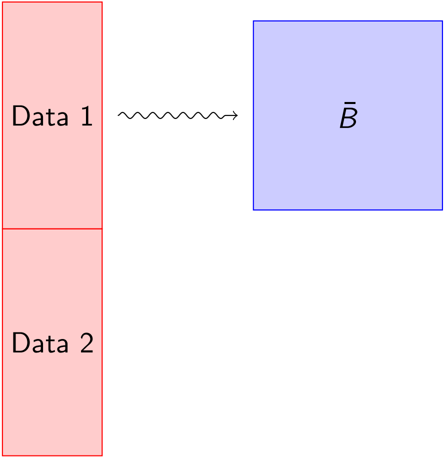
Input: matrix $\bar{B}_{k\times d}$, observations $(X_1, Y_1), \cdots, (X_n, Y_n)$

1: Compute SVD $\bar{B} = UDV^T$.
2: Let V_r^T be the matrix containing r top rows of V^T.
3: Let $\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^d} \sum_{i=1}^n (Y_i - X_i V_r \beta)^2$.
4: Then, output $\widehat{B}_\kappa = (V_r\hat{\beta})^T$.

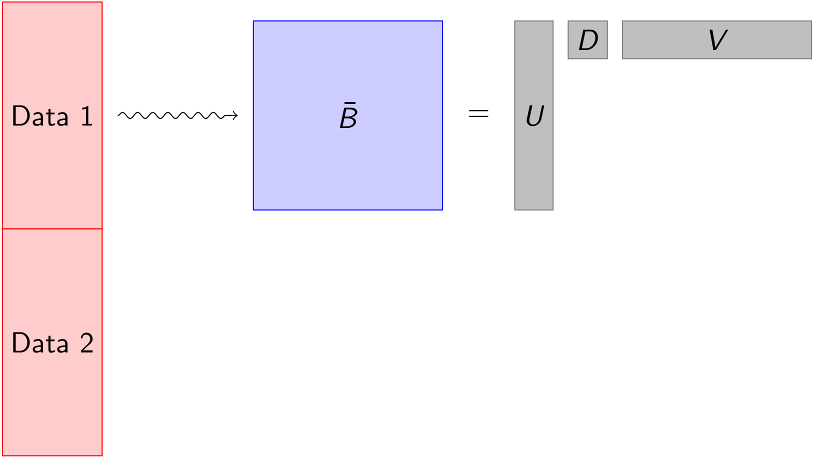
REAL-Estimator



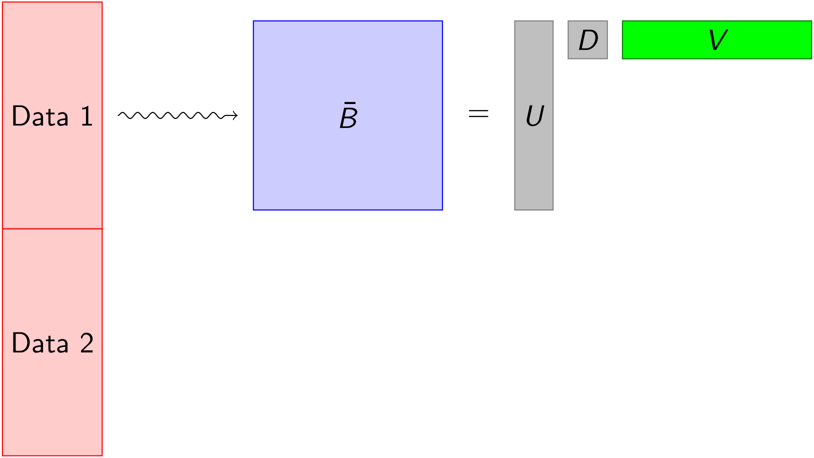
REAL-Estimator



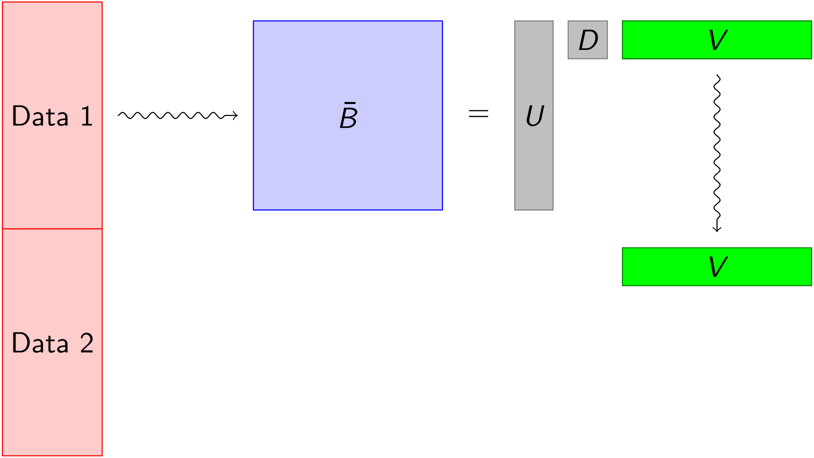
REAL-Estimator



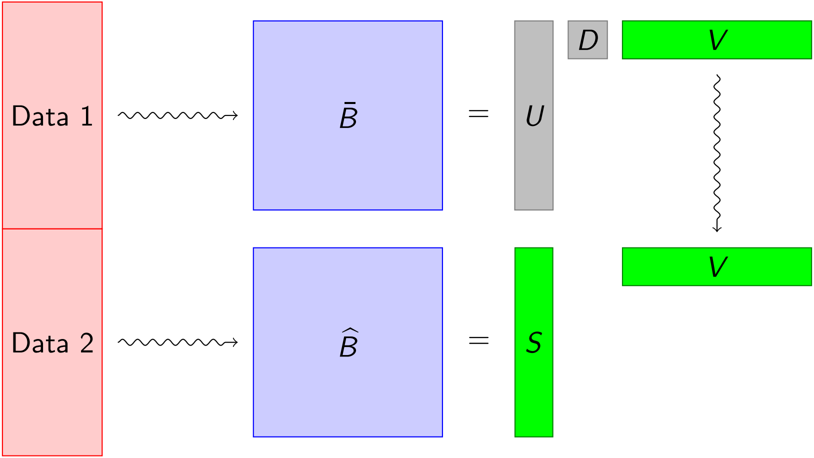
REAL-Estimator



REAL-Estimator

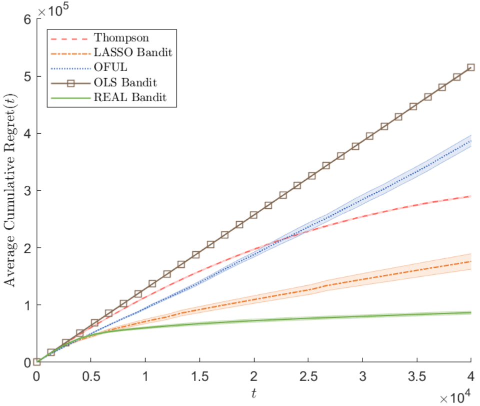


REAL-Estimator



Simulations

- B: 200×201 of rank 3,
- SD of noise (σ): 1,
- Context vectors (X_t): vectors of length 201 with i.i.d. standard normal entries.



References

Tze Leung Lai, and Herbert Robbins
Asymptotically efficient adaptive allocation rules
Advances in applied mathematics 6.1 (1985): 4-22.

Emmanuel J. Cands and Benjamin Recht
Exact matrix completion via convex optimization
Foundations of Computational mathematics 9.6 (2009): 717.

Alexander Goldenshluger and Assaf Zeevi
A linear response bandit problem
Stochastic Systems 3.1 (2013): 230-261.

Hamsa Bastani and Mohsen Bayati
Online decision-making with high-dimensional covariates
(2015).

Sahand Negahban and Martin J. Wainwright
Restricted strong convexity and weighted matrix completion: Optimal bounds with noise
Journal of Machine Learning Research 13.May (2012): 1665-1697.

The End