An Experimental Evaluation of High-Dimensional Multi-Armed Bandits

Naoki Egami Romain Ferrali Kosuke Imai

Princeton University

Talk at Political Data Science Conference
Washington University, St. Louis
May 12, 2016

Political Data Science

- Quantitative Social Science:
  - Causal inference revolution
  - Solve problems by working with governments, NGOs, industries
- Experiments:
  - Multiple treatments and heterogenous treatment effects
  - Sequential experimental design: online experimental platform
- Multi-armed Bandit Experiment:
  - Online learning
  - Select from a large set of treatments
  - Maximize cumulative rewards
  - Applications: election campaigns, conjoint analysis

Detecting Irregularities

- Examples:
  1. Election irregularities (e.g., Ichino and Schündeln 2010; Mebane 2015)
  2. Monitoring government corruption (e.g., Olken 2007)
  3. Tax audit experiment (e.g., Slemrod et al 2001; Kleven et al 2011)

- The Experiment:
  - a large insurance firm processing roadside and heath assistance claims
  - over 100 clerks handle about 1,000 claims each day
  - some claims contain "anomalies"
  - 100 claims are audited every day

  - How to choose 100 claims for audit?
  - Goal: detect and correct as many anomalies as possible
  - Can the bandit algorithm detect more anomalies than experts?

Multi-armed Bandit Problem

- Setting:
  - M treatments or "arms": $\mathcal{Z} = \{z^1, z^2, \cdots z^M\}$
  - sequential sampling indexed by time: $t = 1, 2, \cdots, T$
  - treatment assignment: Z_t
  - potential outcomes: $Y_t(z^m)$
  - observed outcome: $Y_t = Y_t(Z_t)$
- Goal: maximize the cumulative reward $\sum_{t=1}^{T} Y_t$

- Multi-armed bandit algorithm $\rightsquigarrow$ sequential treatment assignment
  1. exploration: try unexplored arms to find a better treatment
  2. exploitation: stay with the currently best performing treatment

Upper Confidence Bound (UCB) Algorithm

- $n_t^m = \sum_{j=1}^t \mathbf{1}\{Z_j = z^m\}$: number of times arm z^m has been assigned
- Sample mean and variance for arm z^m:

$$\hat{\mu}_{t,m} \equiv \frac{1}{n_t^m} \sum_{j=1}^{t} \mathbf{1}\{Z_j = z^m\} Y_j, \quad \hat{\sigma}^2_{t,m} \equiv \frac{1}{n_t^m} \sum_{j=1}^{t} \mathbf{1}\{Z_j = z^m\}(Y_j - \hat{\mu}_{t,m})^2$$

- For the $t+1$st sample, choose:

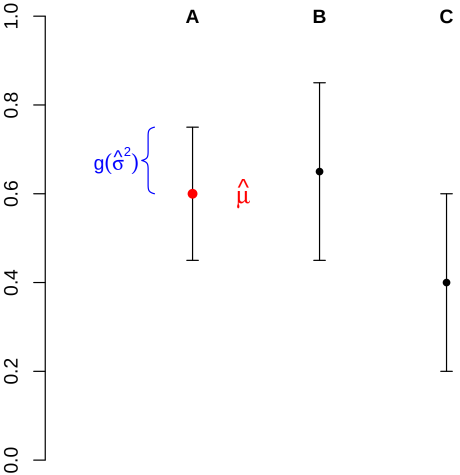
$$Z_{t+1} = \underset{m}{\text{argmax}} \left\{ \underbrace{\hat{\mu}_{t,m}}_{\text{exploitation}} + \underbrace{g(\hat{\sigma}^2_{t,m})}_{\text{exploration}} \right\}$$

- Different algorithm has a different form of $g(\cdot)$

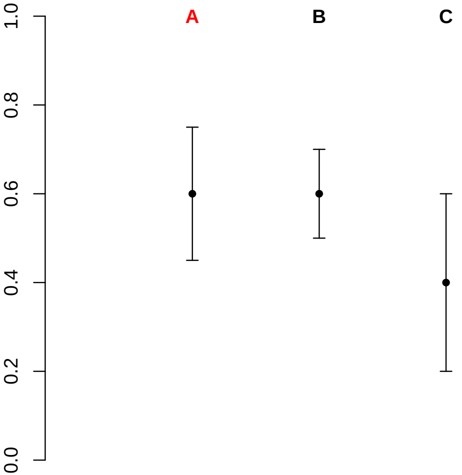
  $\rightsquigarrow$ if $Y_t \mid z^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_m, \sigma_m^2)$, then

$$g(\hat{\sigma}^2_{t,m}) = \sqrt{\hat{\sigma}^2_{t,m} \frac{16 \log(t-1)}{n_m - 1}}$$

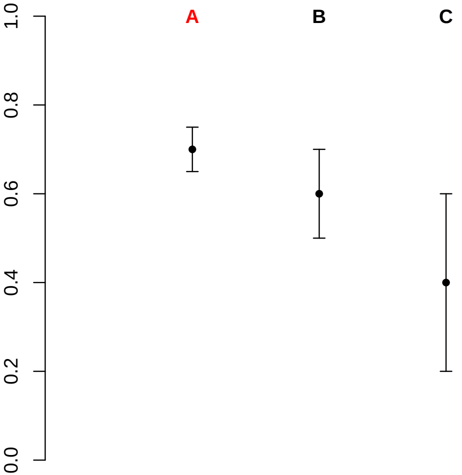
Upper Confidence Bound (UCB) Algorithm



Upper Confidence Bound (UCB) Algorithm



Upper Confidence Bound (UCB) Algorithm



Linear Upper Confidence Bound (Linear UCB) Algorithm

- Motivation: assign multiple treatments at once

- Treatment vector: $Z_t \in \mathcal{Z}$
- Outcome model:
$$\mathbb{E}(Y_t \mid Z_t = z) \;=\; z^\top \beta$$
- Estimate of β at each time t: $\hat{\beta}_t$
- For the $(t+1)$st sample, choose:

$$Z_{t+1} \;=\; \underset{z \in \mathcal{Z}}{\operatorname{argmax}} \, \{z^\top \hat{\beta}_t + g(\widehat{\mathbb{V}(z^\top \hat{\beta}_t)})\}$$

Experimental Evaluation of the Linear UCB Algorithm

- Literature on the multi-armed bandit is largely theoretical
- Many empirical applications in industry
- Few applications published in academic journals

- Experimental comparison between the linear UCB algorithm to experts
- Replication data will be made available for future research

- Expert auditors:
  1. receive about 1,000 claims with their characteristics (only 3 variables!)
  2. choose 20 claims that are "most likely" to contain anomalies
- Linear UCB algorithm:
  1. analyzes the same 1,000 claims with 37 characteristics
  2. selects 20 claims that are "most likely" to contain anomalies
- Each selected claim is examined for anomaly

Linear UCB Algorithm for Anomaly Detection

- Claim characteristics: Z_t
- Binary outcome: $Y_t = 1$ (anomalous), $Y_t = 0$ (otherwise)
- Model:

$$\Pr(Y_t = 1 \mid Z_t = z) \; = \; \text{logit}^{-1}(z^\top \beta)$$

- Estimate β using the logistic ridge regression:

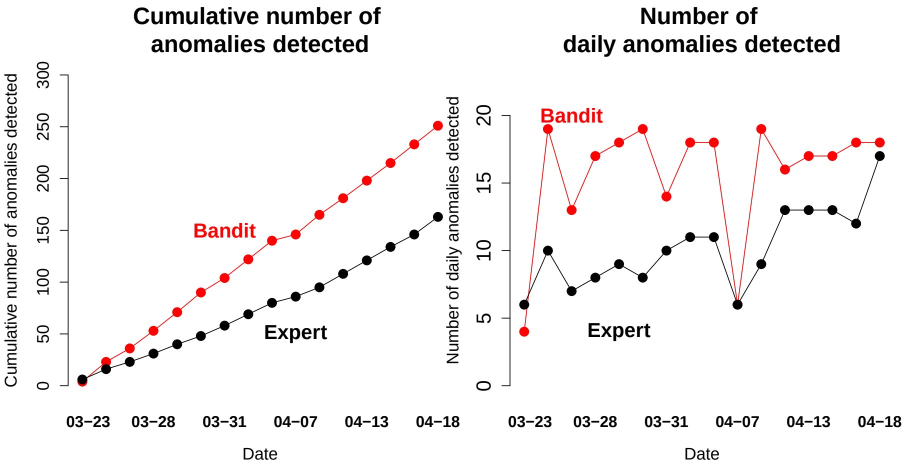
$$\hat{\beta}_t \; = \; \underset{\beta}{\text{argmin}} \sum_{j=1}^{t} \log(1 + \exp\{(1 - 2Y_j)\beta^\top Z_j\}) + \lambda \|\beta\|_2^2$$

  λ is cross-validated with other data
- For each claim at time $t+1$, i.e., $z \in \mathcal{Z}_{t+1}$, compute upper confidence index,

$$p(z) \; = \; \text{logit}^{-1}(z^\top \hat{\beta}_t) + \; \alpha \sqrt{z^\top (\mathbf{Z}^{(t)\top} \mathbf{Z}^{(t)} + \lambda \mathbf{I})^{-1} z}$$

  α is set to 1, which is a typical choice
- Chose 20 claims with the greatest values of $p(z)$

Bandit Beats Experts



High-Dimensional Linear UCB Algorithm

- Extend the Linear UCB algorithm to a high-dimensional setting:
- Our application: variable selection by experts
- What about other variables? Interactions?
  $\rightsquigarrow$ High-dimensional bandit

- Sensitive to the tuning parameter α:

$$p(z) = \text{logit}^{-1}(z^\top \hat{\beta}_t) + \boldsymbol{\alpha}\sqrt{z^\top (\mathbf{Z}^{(t)\top}\mathbf{Z}^{(t)} + \lambda \mathbf{I})^{-1} z}$$

- Cross-validation is too expensive
- Variable selection removes this sensitivity

Simulation Setting

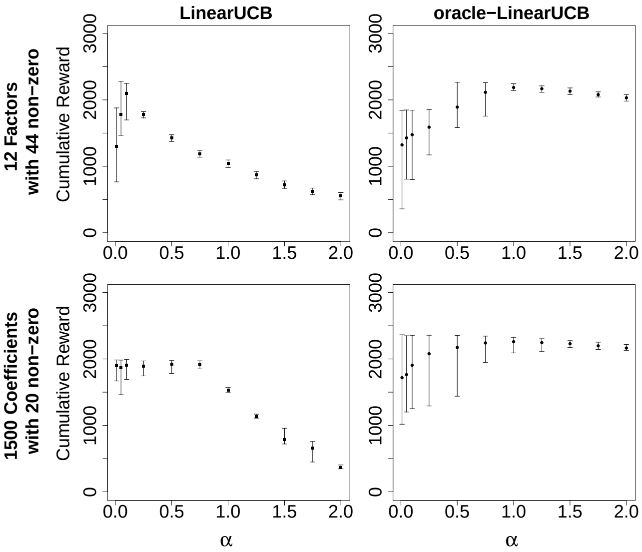
- Goal: Investigate the sensitivity to α

- Outcome model: $\Pr(Y_t = 1 \mid Z_t) = \text{probit}(Z_t^\top \beta)$
- Sample size: $T = 3000$
- Compare 4 bandit algorithms:
  1. Linear UCB (Li *et al.* 2010)
  2. oracle-Linear UCB: known sparsity structure from the start
  3. select-Linear UCB: variable selection at $t = 500$ out of $T = 3,000$
  4. oracle-Linear UCB*: oracle variable selection at $t = 500$
- Change α from 0.01 to 2 following Li *et al.* (2010)
- 100 simulations for each α

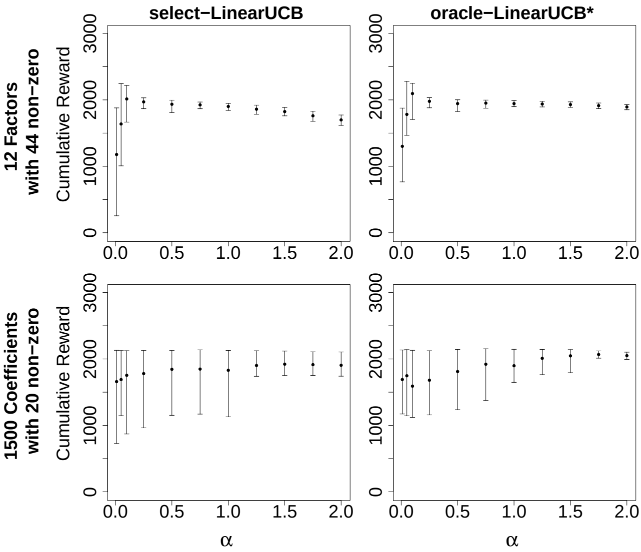
- Simulation 1: Factorial randomized experiments
  - 12 factors, each having 5 levels
  - 3 factors and their two-way interactions are non-zero
  - 44 non-zero coefficients among a total of 1,105 coefficients

- Simulation 2: Independent discrete covariates
  - 1,500 covariates
  - 20 non-zero coefficients out of 1,500 coefficients

Sensitivity of High-Dimensional Linear Bandit



Variable Selection Removes Sensitivity



Theory of Regret Bound

- mean of M arms: $\{\mu^1, \mu^2, \cdots, \mu^M\}$
- mean of the best arm: $\tilde{\mu} = \max_m \mu_m$
- difference in means: $\Delta_m = \tilde{\mu} - \mu_m$
- (Cumulative) regret:

$$R_T \equiv \sum_{t=1}^{T} \sum_{j=1}^{M} \mathbf{1}\{Z_t = z_m\}\Delta_m$$

- Expected regret $\mathbb{E}(R_T)$ of *any* algorithm is bounded below by $o(\log T)$ asymptotically (Lai and Robbins 1985)
- What about the upper bound?
- Example: UCB-Normal (Auer *et al.* 2002)

$$\underbrace{c_1 \log T \sum_{m:\mu_m \neq \tilde{\mu}} \frac{\sigma_m^2}{\Delta_m}}_{\text{exploration}} + \underbrace{(c_2 + 8 \log T) \sum_{m=1}^{M} \Delta_m}_{\text{exploitation}}$$

Regret Bound for the Linear UCB with Variable Selection

- Best treatment: $\tilde{z}$
- Number of coefficients: d
- Number of non-zero coefficients: $s < d$
- Regret : $R_T \equiv \sum_{t=1}^{T} (\tilde{z} - Z_t)^\top \beta$
- maximum instantaneous regret: $\tilde{r} = \max_z (\tilde{z} - z)^\top \beta \leq 2 \max_z |z^\top \beta|$
- T_0: number of observations at the initialization stage
- T_s: timing of variable selection
- Bounds for expected regret:

$$\mathcal{B}(R_T) = \tilde{r}\, T_0 + \underbrace{2\tilde{r} + 2\alpha c_d \sqrt{d} \log^{3/2}(T)\sqrt{T}}_{\text{high dimensional bandit}}$$

$$\mathcal{B}(R_T^{\text{oracle}}) = \tilde{r}\, T_0 + \underbrace{2\tilde{r} + 2\alpha c_s \sqrt{s} \log^{3/2}(T)\sqrt{T}}_{\text{oracle bandit}}$$

$$\mathcal{B}(R_T^{\text{select}}) = \mathcal{B}(R_T^{\text{oracle}}) + \Pr(\text{incorrect selection}) \times \tilde{r}(T - T_s)$$

Sensitivity to the Tuning Parameter

- Result 1: Variable selection lowers the bounds:

$$\mathcal{B}(R_T^{\text{oracle}}) \;\leq\; \mathcal{B}(R_T^{\text{select}}) \;\leq\; \mathcal{B}(R_T)$$

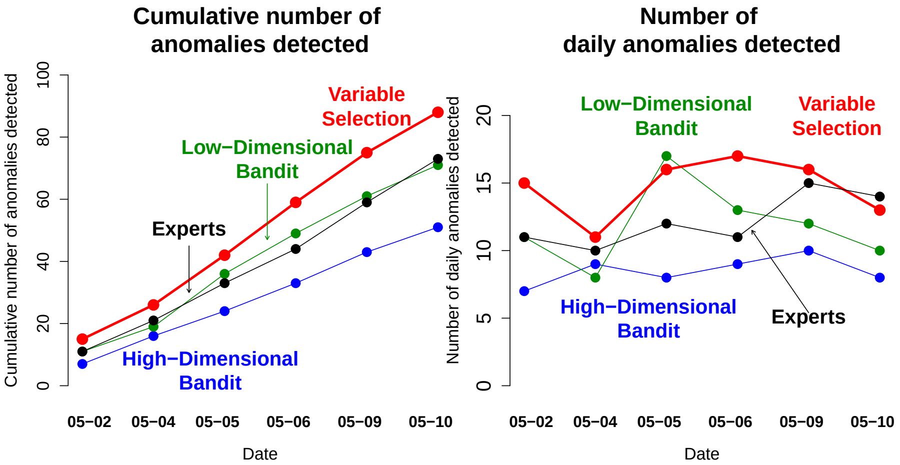
- Result 2: Variable selection reduces the sensitivity to α:

$$\frac{\partial \mathcal{B}(R_T)}{\partial \alpha} \;>\; \frac{\partial \mathcal{B}(R_T^{\text{select}})}{\partial \alpha} \;=\; \frac{\partial \mathcal{B}(R_T^{\text{oracle}})}{\partial \alpha}$$

Experimental Evaluation of High-Dimensional Bandit

- 3 bandit algorithms:
  1. Low-dimensional bandit: 26 variables selected by experts
  2. High-dimensional bandit: all main and 2-way interaction effects of 37 variables
  3. Variable selection bandit: Lasso on High-dimensional bandit everyday
- Procedure of multi-armed bandit algorithm:
  1. each algorithm analyzes the same 1,000 claims
  2. each selects 20 claims that are "most likely" to contain anomalies
  3. all selected claims will be audited
- Expert auditors follow the same protocol as before

Preliminary Results



Conclusion

- Political data science:
  - Causal inference revolution, partnerships with non-academics
  - Causal heterogeneity $\rightsquigarrow$ multiple treatments, online learning
  - Multi-armed bandit experiment
- Experimental evaluation
  - Detecting irregularities
  - Bandit algorithm outperforms experts
  - On-going experiment: high-dimensional bandit
- Theory: benefits of variable selection
  - High-dimensional bandit $\rightsquigarrow$ sensitive to tuning parameter
  - Variable selection removes this sensitivity
- Other applications: election campaign, conjoint analysis