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 $\leadsto$ 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

- \( n^m_t = \sum_{j=1}^{t} 1\{Z_j = z^m\} \): number of times arm \( z^m \) has been assigned
- Sample mean and variance for arm \( z^m \):
  \[
  \hat{\mu}_{t,m} \equiv \frac{1}{n^m_t} \sum_{j=1}^{t} 1\{Z_j = z^m\} Y_j, \quad \hat{\sigma}^2_{t,m} \equiv \frac{1}{n^m_t} \sum_{j=1}^{t} 1\{Z_j = z^m\} (Y_j - \hat{\mu}_{t,m})^2
  \]
- For the \( t+1 \)st sample, choose:
  \[
  Z_{t+1} = \arg\max_m \left\{ \hat{\mu}_{t,m} + g(\hat{\sigma}^2_{t,m}) \right\}
  \]
- Different algorithm has a different form of \( g(\cdot) \)
  \( \sim \) if \( Y_t \mid z^m \) i.i.d. \( \mathcal{N}(\mu_m, \sigma^2_m) \), 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

\[ g(\hat{\sigma}^2) \]

\[ \hat{\mu} \]

A  B  C
Upper Confidence Bound (UCB) Algorithm

Egami, Ferrali and Imai (Princeton)
Multi-armed Bandit
Wash U. (May 12, 2016)
Upper Confidence Bound (UCB) Algorithm

A  B  C

Egami, Ferrali and Imai (Princeton)
Linear Upper Confidence Bound (Linear UCB) Algorithm

- Motivation: assign multiple treatments at once

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

\[
Z_{t+1} = \arg\max_{z \in \mathcal{Z}} \{ z^\top \hat{\beta}_t + g(\sqrt{\mathbb{V}(z^\top \hat{\beta}_t)}) \}\]
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
Claim characteristics: $Z_t$

Binary outcome: $Y_t = 1$ (anomalous), $Y_t = 0$ (otherwise)

Model:

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

Estimate $\beta$ using the logistic ridge regression:

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

$\lambda$ is cross-validated with other data

For each claim at time $t + 1$, i.e., $z \in Z_{t+1}$, compute upper confidence index,

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

$\alpha$ is set to 1, which is a typical choice

Chose 20 claims with the greatest values of $p(z)$
Bandit Beats Experts

Cumulative number of anomalies detected

Number of daily anomalies detected

Cumulative number of anomalies detected

Date

03–23 03–28 03–31 04–07 04–13 04–18

Number of daily anomalies detected

0 5 10 15 20

Date

03–23 03–28 03–31 04–07 04–13 04–18

Egami, Ferrali and Imai (Princeton)  Multi-armed Bandit

Wash U. (May 12, 2016)
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?
  \[ \sim \text{High-dimensional bandit} \]

- Sensitive to the tuning parameter \( \alpha \):
  \[ p(z) = \logit^{-1}(z^\top \hat{\beta}_t) + \alpha \sqrt{z^\top (Z(t)^\top Z(t) + \lambda I)^{-1} z} \]

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

- **Goal**: Investigate the sensitivity to $\alpha$

- **Outcome model**: $\Pr(Y_t = 1 \mid Z_t) = \text{probit}(Z_t^T \beta)$

- **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 $\alpha$ from 0.01 to 2 following Li *et al.* (2010)**

- **100 simulations for each $\alpha$**
• **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

Egami, Ferrali and Imai (Princeton)  Multi-armed Bandit  Wash U. (May 12, 2016)  16 / 23
Variable Selection Removes Sensitivity

12 Factors with 44 non-zero
1500 Coefficients with 20 non-zero

Cumulative Reward

select–LinearUCB

oracle–LinearUCB*

\(\alpha\)

Egami, Ferrali and Imai (Princeton)  Multi-armed Bandit  Wash U. (May 12, 2016)  17 / 23
Theory of Regret Bound

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

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

- 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)

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

exploration

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:

$$B(R_T) = \tilde{r} T_0 + 2\tilde{r} + 2\alpha c d \sqrt{d \log^{3/2}(T) \sqrt{T}}$$

high dimensional bandit

$$B(R_{T}^{\text{oracle}}) = \tilde{r} T_0 + 2\tilde{r} + 2\alpha c s \sqrt{s \log^{3/2}(T) \sqrt{T}}$$

oracle bandit

$$B(R_{T}^{\text{select}}) = B(R_{T}^{\text{oracle}}) + \Pr(\text{incorrect selection}) \times \tilde{r}(T - T_s)$$
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 \( \alpha \):
  \[ \frac{\partial \mathcal{B}(R_T)}{\partial \alpha} > \frac{\partial \mathcal{B}(R_T^{\text{select}})}{\partial \alpha} = \frac{\partial \mathcal{B}(R_T^{\text{oracle}})}{\partial \alpha} \]
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

Cumulative number of anomalies detected

Number of daily anomalies detected

Egami, Ferrali and Imai (Princeton)  Multi-armed Bandit  Wash U. (May 12, 2016)
Conclusion

- Political data science:
  - Causal inference revolution, partnerships with non-academics
  - Causal heterogeneity $\Rightarrow$ 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 $\Rightarrow$ sensitive to tuning parameter
  - Variable selection removes this sensitivity

- Other applications: election campaign, conjoint analysis