Statistical Inference with M-Estimators on Adaptively Collected Data (Kelly Zhang et al, 2021)

DSC 242: High-Dimensional Probability and Statistics (Gabriel Riegner, 2024)

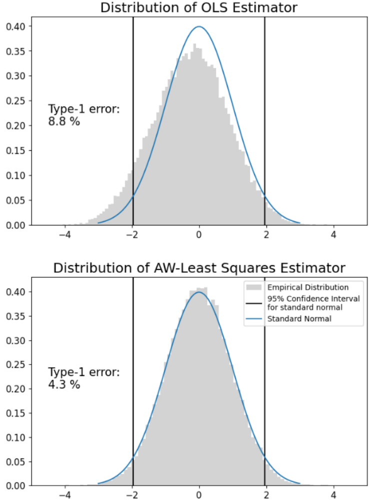
Setting

Methodology

Assumptions

Numerical Analysis

Conclusions



Introduction

Bandit Algorithms:

- **Exploration:** gather new information
- **Exploitation:** use existing information for decisions

Objectives:

- **Regret Minimization:** select choice that would cause least 'regret' (within experiment)
- **Statistical Inference:** hypothesis testing to inform future experiments (between experiments)

Challenges:

- **Sampling:** adaptive sampling introduces time dependence on observed data (not independent nor identically distributed)
- **Type-I errors:** leads to unstable M-estimator variances and inflated false positive rates

Notation

Data Structures:

- Contexts: $\{X_t\}_{t=1}^T$
- Actions: $\{A_t\}_{t=1}^T$
- Rewards: $R_t = f(Y_t)$, deterministic function of outcomes

Outcomes:

- Primary: $\{Y_t\}_{t=1}^T$, theoretical outcome of interest
- Potential: $\{Y_t(a) : a \in \mathcal{A}\}$, theoretical space for all actions
- Observed: $Y_t = Y_t(A_t)$, realization of a specific action

Definitions:

- Model: $\theta^*(\mathcal{P}) \in \underset{\theta\in\Theta}{\operatorname{argmax}}\, \mathbb{E}_{\mathcal{P}}[m_\theta(Y_t, X_t, A_t)|X_t, A_t]$
- Estimator: $\hat{\theta}_T = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \frac{1}{T}\sum_{t=1}^T m_\theta(Y_t, X_t, A_t)$

  m_θ is the criterion function, e.g. squared loss or log-likelihood

Objective:

- Derive asymptotically valid confidence regions for $\theta^*(\mathcal{P})$

Adaptively Weighted M-Estimators

Motivation:

- Adaptive policies introduce dependencies that lead to unstable standard errors and false positives
- Adaptive weights stabilize standard errors

Weighted M-estimator:

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{T} \sum_{t=1}^{T} W_t m_\theta(Y_t, X_t, A_t)$$

Square-root Importance Weights:

$$W_t = \sqrt{\frac{\pi_t^{\mathrm{sta}}(A_t, X_t)}{\pi_t(A_t \mid X_t, \mathcal{H}_{t-1})}}$$

- Stabilizing policy: $\pi_t^{\mathrm{sta}}(A_t, X_t) = 1/|\mathcal{A}|$
- Adaptive policy: $\pi_t(A_t, X_t, \mathcal{H}_{t-1}) := \mathbb{P}(A_t|\mathcal{H}_{t-1}, X_t)$

  $\mathcal{H}_{t-1} = \{(X_{t'}, A_{t'}, Y_{t'})\}_{t'=1}^{t-1}$ are the histories

Weighted M-Estimator Properties

- **Consistency:**

$$\hat{\theta}_T \xrightarrow{p} \theta^*(\mathcal{P}) \text{ uniformly over } \mathcal{P} \in \mathbf{P}$$

- **Asymptotic Normality:**

$$\sqrt{T}(\hat{\theta}_T - \theta^*(\mathcal{P})) \xrightarrow{D} \mathcal{N}\left(0, \ddot{M}_T(\hat{\theta}_T)^{-1}\Sigma_T(\mathcal{P})\ddot{M}_T(\hat{\theta}_T)^{-1}\right)$$

- **Valid Confidence Regions:**

$$\liminf_{T\to\infty} \inf_{\mathcal{P}\in\mathbf{P}} \mathbb{P}(\theta^*(\mathcal{P}) \in C_T(\alpha)) \geq 1 - \alpha$$

$\ddot{M}_T(\theta)$ is the Hessian of the criterion function

$\Sigma_T(\mathcal{P})$ is the covariance matrix of the criterion under the pre-specified stabilizing policy $\{\pi_t^{\mathrm{sta}}\}_{t=1}^T$

Assumptions

Conditions for adaptively weighted M-estimators to be **asymptotically normal:**

- **Stochastic Bandit Environment:** $\{X_t, Y_t(a) : a \in \mathcal{A}\} \overset{i.i.d.}{\sim} \mathcal{P}$ over $t \in [1 : T]$.
- **Differentiable:** the criterion function is differentiable for all $\theta \in \Theta$ and (x, y) in the joint support.
- **Lipschitz:** $|m_\theta - m_{\theta'}| \leq g\|\theta - \theta'\|_2$, with bounded $\mathbb{E}[g^2]$.
- **Bounded Space:** parameter space Θ is bounded and open.
- **Moments:** fourth moments of m_θ, $\dot{m}_\theta$, and $\ddot{m}_\theta$ are uniformly bounded.
- **Bounded Importance Ratios:** $0 < \rho_{\min} \leq \frac{\pi^{\text{sta}}}{\pi} \leq \rho_{\max} < \infty$.

Simulation Setting

Setting:

- Two-arm contextual bandit: $\mathcal{A} = \{1, 2\}$
- Contexts X_t, actions A_t, outcomes Y_t

$\mathbb{E}_{\mathcal{P}}[Y_t | X_t, A_t = a] = X_t^\top \theta_a^*(\mathcal{P})$ w.p. 1

Least Squares (LS) Estimators:

- **Unweighted/Ordinary LS:**

$$\hat{\theta}_{T,a}^{\mathrm{OLS}} := \arg\min_{\theta_a} \sum_{t=1}^{T} \mathbb{1}_{A_t=a} \big(Y_t - X_t^\top \theta_a \big)^2$$

- **Adaptively Weighted LS:**

$$\hat{\theta}_{T,a}^{\mathrm{AW-LS}} := \arg\min_{\theta_a} \sum_{t=1}^{T} W_t \mathbb{1}_{A_t=a} \left(Y_t - X_t^\top \theta_a \right)^2$$

$W_t = \sqrt{\frac{\pi^{\mathrm{sta}}(A_t, X_t)}{\pi_t(A_t | X_t, \mathcal{H}_{t-1})}}$, where $\pi^{\mathrm{sta}}(a, x) = 1/2$

Zhang et al Results

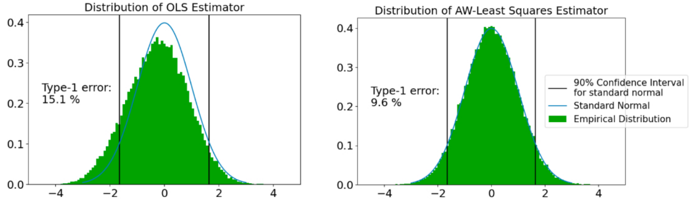
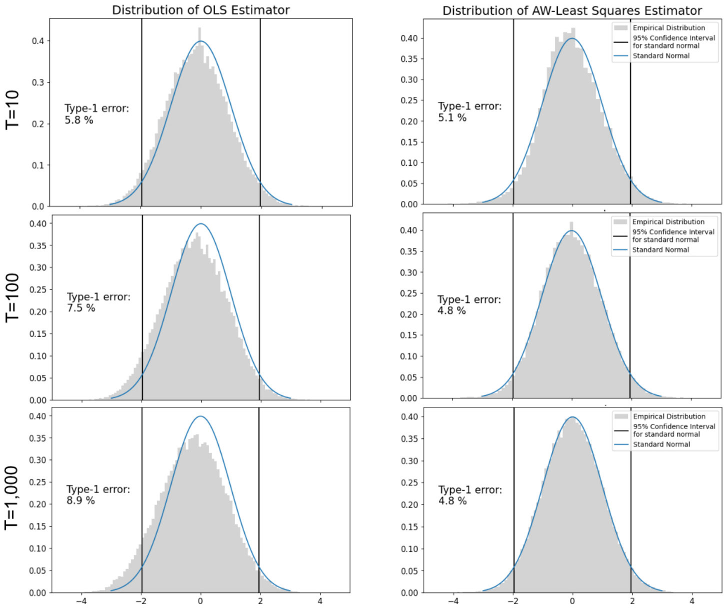


Figure 1: The empirical distributions of the weighted and unweighted least-squares estimators for $\theta_1^*(\mathcal{P}) := \mathbb{E}_{\mathcal{P}}[Y_t(1)]$ in a two arm bandit setting where $\mathbb{E}_{\mathcal{P}}[Y_t(1)] = \mathbb{E}_{\mathcal{P}}[Y_t(0)] = 0$. We perform Thompson Sampling with $\mathcal{N}(0,1)$ priors, $\mathcal{N}(0,1)$ errors, and $T = 1000$. Specifically, we plot $\sqrt{\sum_{t=1}^T A_t}(\hat{\theta}_{T,1}^{\text{OLS}} - \theta_1^*(\mathcal{P}))$ on the left and $\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \sqrt{\frac{0.5}{\pi_t(1)}} A_t\right) (\hat{\theta}_{T,1}^{\text{AW-LS}} - \theta_1^*(\mathcal{P}))$ on the right.

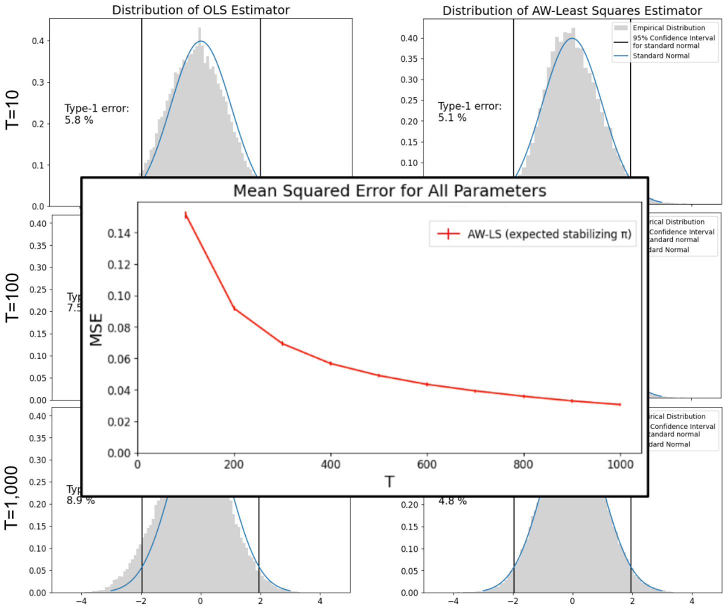
My Results



$$\sqrt{\sum_{t=1}^{T} A_t}\,(\hat{\theta}_{T,1}^{\text{OLS}} - \theta_1^*(\mathcal{P}))$$

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sqrt{\frac{0.5}{\pi_t(1)}} A_t\right) (\hat{\theta}_{T,1}^{\text{AW-LS}} - \theta_1^*(\mathcal{P}))$$

My Results



$$\sqrt{\sum_{t=1}^{T} A_t}\,(\hat{\theta}^{\text{OLS}}_{T,1} - \theta^*_1(\mathcal{P}))$$

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{\frac{0.5}{\pi_t(1)}}A_t\right)(\hat{\theta}^{\text{AW-LS}}_{T,1} - \theta^*_1(\mathcal{P}))$$

Conclusions

- Bandit algorithms introduce data dependencies that make traditional statistical inference unreliable
- Adaptively weighted M-estimators, using square-root importance weights, provide valid statistical inference from data collected with bandit algorithms
- Least squares simulations confirm theoretical asymptotic results, that this approach stabilizes standard errors and maintains valid Type-I error rates