Causal Inference with Noisy and Missing Covariates

Vikram Mullachery

Feb 2019

References

- **Causal Inference with Noisy and Missing Covariates** by Nathan Kallus, Xiaojie Mao, Madeleine Udell (2018)

Causal Inference from observational data

Summary

- Observational vs. experimental
- Samples contain Treated $T_i = 1$ and Control $T_i = 0$
- Covariates U_i (either observed or unobserved)
- Outcome of interest Y_i
- Confounders are covariates that effect both the outcome and treatment

Causal Inference from observational data

Assumptions

- Stable Unit Treatment Value Assignment
- Consistency
- Ignorability
- Positivity

Causal Inference from observational data

To estimate

- Average Treament Effect, ATE: $\mathbb{E}[Y^1] - \mathbb{E}[Y^0]$

Causal Inference from observational data

Potential Outcome framework

- Matching (Mahalanobis-distance, calipers etc.)
- Propensity score based inverse probability weighting (IPW)
- Doubly Robust methods

Techniques

- Assume a causal relationship (a causal graph)
- Identify set of confounders to be controlled (or measured)
- Use backdoor criteria or frontdoor criteria etc. to assist discovery of valid adjustment set

Problem Setting

Notation

- Treatment $T \in \{0, 1\}^N$
- Unobserved Covariates $U \in \mathbb{R}^{N \times r}$
- Outcome $Y \in \mathbb{R}^{N \times 1}$
- Observed noisy and missing covariates $X \in \mathbb{R}^{N \times p}$

Assumptions

- Linear $Y_i = U_i\alpha + \tau T_i + \epsilon_i$
- Low rank matrix factorization of the observed X to yield confounders $X = UV^T + W$
- Exponential Family Matrix Completion preprocessing

Contribution

Claim

- Preprocessing to augment wide variety of causal inferences
- Matrix factorization based preprocessing is a general framework
- Seamlessly integrated into regression adjustment, propensity score reweighting, matching etc.
- Bounded error on the induced average treatment effect estimator

Inference with low rank matrix decomposition

Problem Formulation

- Estimate $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$
- **Unconfoundedness assumption** $\mathbb{P}(Y_i | T_i, U_i) = \mathbb{P}(Y_i | U_i) \qquad \forall i$
- $X \in \mathbb{R}^{N \times p}$ and observed over $\Omega \subset [N] \times [p], p < N$
- Generative assumption $X_{ij} \sim \mathcal{N}(U_i^T V_j, 1)$
- $W \in \mathbb{R}^{N \times p}$ independent with (mean, variance) $= (0, \sigma_w^2)$
- Linear $Y_i = U_i^T \alpha + \tau T_i + \epsilon_i$
- Additive noise model $X = UV^T + W$

Inference with low rank matrix decomposition

Measurement Noise and Bias

Asymptotic bias of least squares estimator in linear regression of Y_i on X_i, T_i

$$\frac{\mathbb{E}(T_i U_i)\mathbb{E}(U_i^T U_i)^{-1}[\frac{1}{\sigma_w^2} V^T V + \mathbb{E}(U_i^T U_i)^{-1}]^{-1}\alpha}{\mathbb{E}(T_i^2) - \mathbb{E}(T_i U_i)[\frac{1}{\sigma_w^2} V^T V + \mathbb{E}(U_i^T U_i)]^{-1}\mathbb{E}(U_i^T U_i)}$$

This asymptotically diminishes to 0, when $||V|| \to \infty$

Inference with low rank matrix decomposition

Low rank matrix factorization preprocessing

Low rank Assumption Observed X is a noisy realization of a low rank matrix $\Phi \in \mathbb{R}^{N \times p}$ with rank $r \ll \min\{N, p\}$

Missing Completely At Random Assumption
$\forall i, j \in \Omega, i \sim \text{Unif}([N]), j \sim \text{Unif}([p])$

Natural Exponential Family Assumption
$\mathbb{P}(X_{ij}|\Phi_{ij}) = h(X_{ij})exp(X_{ij}\Phi_{ij} - G(\Phi_{ij}))$, where $G : \mathbb{R} \mapsto \mathbb{R}$ is the log-partition and strictly convex, $\nabla^2 G \geq e^{-\eta|u|}$, for $\eta > 0$, and $u \in \mathbb{R}$

EFMC estimates using regularized M-estimator:
$\hat{\Phi} = min \frac{-Np}{|\Omega|}[\sum_{(i,j)\in\Omega} \log \mathbb{P}(X_{ij}|\Phi_{ij})] + \lambda||\Phi||_*$
Left singular matrix of $\hat{\Phi}$ is an estimate of the confounder U

Inference with low rank matrix decomposition

Theoretical guarantee (sufficient conditions)

Definition Principal angle between column spaces of two matrices $M, \hat{M}$ is defined as $\angle(\hat{M}, M) = \sqrt{1 - \sigma^2_{\min r,k}(\hat{M}^T M)}$

Theorem There exists a constant $c > 0$ such that with probability at least $1 - 2exp(-c\sqrt{N})$,
$|\hat{\tau} - \tau^*| \leq \frac{\frac{2A}{\sqrt{N}}||T||(\frac{1}{\sqrt{Nr}}||U||)(r\angle(U,\hat{U})) - \frac{\sigma}{N^{1/4}}}{\frac{1}{N}T^T(I-P_U)T - \frac{2}{N}||T||^2\angle(U,\hat{U})}$, which $\rightarrow 0$ as $N \rightarrow \infty$, where $||\alpha||_{max} \leq A$

Inference with low rank matrix decomposition

Theoretical guarantee (sufficient conditions)

- $||\alpha||_{max} \leq A$
- $\frac{1}{\sqrt{Nr}}||U||$
- $\frac{1}{N}T^T(I - P_U)T$ is bounded away from 0
- $r\angle\hat{U}, U \rightarrow 0$ as $N \rightarrow 0$
- Unconfoundedness

Inference with low rank matrix decomposition

Theoretical guarantee(Confounders and Covariates Loadings

- $\underline{v}$, $\bar{v}, c_V, c_L > 0$
- For $i \in [N]$, U_i are i.i.d Gaussian samples with covariance $\Sigma_{r \times r} = LL^T$, full rank $L \in \mathbb{R}^{r \times r}$, such that $\frac{1}{\sqrt{r}}||L|| < c_L$ (Gaussian random design)
- $\underline{v}$p$\leq \sigma_r^2(VL^T) \leq \sigma_1^2(VL^T) \leq \bar{v}p$
- $\frac{\max_j ||V_j||}{||V||_F} \leq \frac{c_V}{\sqrt{p}}, j \in [p]$ (excludes degenerate case)

Inference with low rank matrix decomposition

Theoretical guarantee

- Let X_{ij} be sub-exponential on U_i with parameter σ' for $\forall i, j$
- T_i is almost certainly not a linear combination of U_i
- Suppose EFMC is used as the preprocessing step with $\lambda = 2c_0\sigma'\sqrt{Np}\sqrt{\frac{r\bar{N}\log\bar{N}}{|\Omega|}}$, where $\bar{N} = \max(N, p)$, and $|\Omega| > c_1 r\bar{N}\log\bar{N}$ for $c_0, c_1 > 0$
- $\exists\delta$, s.t. $p^{1+\delta}/N \to 0$

Inference with low rank matrix decomposition

Theoretical guarantee (Consistent Estimator)

Theorem There exist constants $c_2, c_3, c_{\sigma',\eta}$ such that with probability at least $1 - c_2 exp(-c_3 N^{1/2}) - c_2 N^{-1/2} - 2exp(-c_3 p^{\delta})$,

$|\hat{\tau} - \tau| \leq \frac{A c_L c_{\sigma',\eta} c_V \sqrt{\frac{r^5 \bar{r} \bar{N} \log \bar{N}}{|\Omega|}} - \frac{\sigma}{N^{1/4}} [\sqrt{\frac{\underline{v}}{\underline{v}+2\bar{v}}} - \Lambda(r,\bar{N},|\Omega|)]}{[\sqrt{\frac{\underline{v}}{\underline{v}+2\bar{v}}} - \Lambda(r,\bar{N},|\Omega|)][\frac{1}{N} T^T (I - P_U) T - 2\Lambda(r,\bar{N},|\Omega|)]}$, where

$\Lambda(r, \bar{N}, |\Omega|) = c_{\sigma',\eta} c_V \sqrt{\frac{\bar{r} r^3 \bar{N} \log \bar{N}}{|\Omega|}}$, and $\bar{r} = \max r, \log \bar{N}$

Inference with low rank matrix decomposition

Numerical Results (Consistent Estimator)

- $r^5 \bar{r} \bar{N} \log \bar{N} / |\Omega| \to 0$ vs. $r \bar{N} \log \bar{N} / |\Omega| \to 0$
- Numerical results: $N = 1500, p = 1450, r = 5$, that is $r^6 \gg N$

Questions

Thank you