因果推断文献综述

罗姗姗

北京工商大学 数学与统计学院

January, 2025

目录

随机化试验

Randomized Controlled Trial

随机化试验下相关问题 I

- Interference, SUTVA 假设不成立时

Hudgens, M. G., Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, **103**(482), 832–842.

Tchetgen, E. J. T., Van der Weele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, **21**, 55–75.

Liu, L., Hudgens, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association*, **109**(505), 288–301.

Liu, L., Hudgens, M. G., Becker-Dreps, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika*, **103**(4), 829–842.

随机化试验下相关问题 II

Athey, S., Eckles, D., Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, **113**(521), 230–240.

Park, C., Kang, H. (2021). Assumption-Lean Analysis of Cluster Randomized Trials in Infectious Diseases for Intent-to-Treat Effects and Network Effects. *Journal of the American Statistical Association*, **118**, 1195–1206.

- Noncompliance, 非依从

Imbens, G. W., Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62**, 467–475.

Angrist, J. D., Imbens, G. W., Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91**, 444–455.

随机化试验下相关问题 III

Imai, K., Jiang, Z., Malani, A. (2020). Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association*, **116**(534), 632–644.

- Covariate adjustment in RCT

Zhao, A., Ding, P. (2022). To Adjust or not to Adjust? Estimating the Average Treatment Effect in Randomized Experiments with Missing Covariates. *Journal of the American Statistical Association*, **119**(545), 450–460.

Ding, P. (2024). *A first course in causal inference*. CRC Press. (Chapters 3-9)

- Rerandomization

Ding, P. (2024). *A first course in causal inference*. CRC Press. (Chapters 3-9)

观察性研究

Observational Studies

观察性研究中的相关问题 I

- 基础知识

Ding, P. (2024). *A first course in causal inference*. CRC Press. (Chapters 10-15)

- 稳健估计

Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–866.

Lunceford, J. K., Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, **23**, 2937–2960.

Bang, H., Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, **61**, 962–973

观察性研究中的相关问题 II

- Kang, J. D. Y., Schafer, J. L. (2007). Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, **22**, 523–539.
- Cao, W., Tsiatis, A. A., Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, **96**(3), 723–734.
- Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, **97**, 661–682.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, **21**(1), C1–C68.

观察性研究中的相关问题 III

Pan, Y., Zhao, Y. Q. (2020). Improved doubly robust estimation in learning optimal individualized treatment rules. *Journal of the American Statistical Association*, **116**(533), 283–294.

- extreme propensity score

Yang, S., Ding, P. (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika*, **105**(2), 487–493.

Li, F., Morgan, K. L., Zaslavsky, A. M. (2017). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, **113**(521), 390–400.

- 迁移学习，数据融合

观察性研究中的相关问题 IV

Yang, S., Kim, J. K., Song, R. (2020). Doubly robust inference when combining probability and non-probability samples with high dimensional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **82**(2), 445–465.

Yang, S., Ding, P. (2020). Combining multiple observational data sources to estimate causal effects. *Journal of the American Statistical Association*, **115**(531), 1540–1554.

Li, X., Miao, W., Lu, F., Zhou, X.-H. (2023). Improving efficiency of inference in clinical trials with external control data. *Biometrics*, **79**(1), 394–403.

Wu, L., Yang, S. (2023). Transfer learning of individualized treatment rules from experimental to real-world data. *Journal of Computational and Graphical Statistics*, **32**(3), 1036–1045.

Chu, J., Lu, W., Yang, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*, **110**(4), 913–931.

观察性研究中的相关问题 V

- Kallus, N., Mao, X. (2024). On the role of surrogates in the efficient estimation of treatment effects with limited outcome data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, to appear.
- Wu, P., Luo, S., Geng, Z. (2024). On the comparative analysis of average treatment effects estimation via data combination. *Journal of the American Statistical Association*, to appear.

- Optimal treatment regime (OTR); Policy learning
  - Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B*, **65**, 331–355. (A-learning)
  - Qian, M., Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, **39**, 1180. (Q-learning)

观察性研究中的相关问题 VI

Zhang, Y., Laber, E. B., Tsiatis, A., Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, **71**, 895–904.

Zhao, Y.-Q., Zeng, D., Rush, A. J., Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, **107**, 1106–1118.

Zhao, Y.-Q., Zeng, D., Laber, E. B., Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, **110**, 583–598.

Wang, Y., Fu, H., Zeng, D. (2018). Learning Optimal Personalized Treatment Rules in Consideration of Benefit and Risk: With an Application to Treating Type 2 Diabetes Patients With Insulin Therapies. *Journal of the American Statistical Association*, 113(521), 1–13.

观察性研究中的相关问题 VII

- Chu, J., Lu, W., Yang, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*, **110**(4), 913–931.
- Li, C., Zeng, D., Zhu, W. (2024). A robust covariate-balancing method for learning optimal individualized treatment regimes. *Biometrika*, to appear.
- Pan, Y., Zhao, Y. Q. (2020). Improved doubly robust estimation in learning optimal individualized treatment rules. *Journal of the American Statistical Association*, **116**(533), 283–294.
- Guo, W., Zhou, X. H., Ma, S. (2021). Estimation of optimal individualized treatment rules using a covariate-specific treatment effect curve with high-dimensional covariates. *Journal of the American Statistical Association*, **116**(533), 309–321.

工具变量

Instrumental Variable

工具变量的相关问题 I

- 基础知识

Ding, P. (2024). *A first course in causal inference*. CRC Press. (Chapters 21-25)

- 非参平均因果作用识别

Wang, L., Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**(3), 531–550.

Cui, Y., Tchetgen Tchetgen, E. (2020). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, **116**(533), 162–173.

Sun, B., Cui, Y., Tchetgen Tchetgen, E. (2022). Selective machine learning of the average treatment effect with an invalid instrumental variable. *Journal of Machine Learning Research*, **23**(204), 1–40.

工具变量的相关问题 II

- Ai, C., Huang, L., Zhang, Z. (2022). A simple and efficient estimation of average treatment effects in models with unmeasured confounders. *Statistica Sinica*, **32**(2), 1007–1026. （nonparameteric estimation）

- Optimal treatment regime (OTR); Policy learning
  - Cui, Y., Tchetgen Tchetgen, E. (2020). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, **116**(533), 162–173.
  - Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C., Luedtke, A. (2021). Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*, **116**(533), 174–191.

工具变量的相关问题 III

- 高维（无效）工具变量

Lin, W., Feng, R., Li, H. (2015). Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. *Journal of the American Statistical Association*, **110**(509), 270–288.

Kang, H., Zhang, A., Cai, T. T., Small, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American Statistical Association*, **111**(513), 132–144.

Guo, Z., Kang, H., Cai, T. T., Small, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **80**(4), 793–815.

工具变量的相关问题 IV

- Windmeijer, F., Farbmacher, H., Davies, N., Davey Smith, G. (2019). On the use of the lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association*, **114**(527), 1339–1350.
- Windmeijer, F., Liang, X., Hartwig, F. P., Bowden, J. (2021). The confidence interval method for selecting valid instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **83**(4), 752–776.

- 无效工具变量

- Sun, B., Cui, Y., Tchetgen Tchetgen, E. (2022). Selective machine learning of the average treatment effect with an invalid instrumental variable. *Journal of Machine Learning Research*, **23**(204), 1–40.

工具变量的相关问题 V

- Liu, Z., Ye, T., Sun, B., Schooling, M., Tchetgen, E. T. (2023). Mendelian randomization mixed-scale treatment effect robust identification and estimation for causal inference. *Biometrics*, **79**, 2208–2219.
- Dukes, O., Richardson, D. B., Shahn, Z., Robins, J. M., Tchetgen Tchetgen, E. J. (2024). Using negative controls to identify causal effects with invalid instrumental variables. *Biometrika*, to appear.

- 个体因果作用

- Vuong, Q., Xu, H. (2017). Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics*, **8**(2), 589–610.

工具变量的相关问题 VI

Feng, Q., Vuong, Q., Xu, H. (2020). Estimation of heterogeneous individual treatment effects with endogenous treatments. *Journal of the American Statistical Association*, **115**(529), 231–240.

- 迁移学习；两样本孟德尔随机化

Sun, B., Miao, W. (2022). On semiparametric instrumental variable estimation of average treatment effects through data fusion. *Statistica Sinica*, **32**, 569–590.

Zhao, Q., Wang, J., Spiller, W., Bowden, J., Small, D. S. (2019). Two-sample instrumental variable analyses using heterogeneous samples. *Statistical Science*, **34**(2), 317–333.

Zhao, Q., Wang, J., Hemani, G., Bowden, J., Small, D. S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics*, **48**, 1742–1769.

工具变量的相关问题 VII

Shuai, K., Luo, S., Li, W., He, Y. (2024). Identifying causal effects using instrumental variables from the auxiliary population. *Statistica Sinica*, To appear.

- 缺失数据

Chen, H., Geng, Z., Zhou, X. H. (2009). Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data. *Biometrics*, **65**, 675–682.

- 孟德尔随机化

VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M., Kraft, P. (2014). Methodological challenges in Mendelian randomization. *Epidemiology*, **25**, 427.

Bowden, J., Davey Smith, G., Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, **44**(2), 512.

工具变量的相关问题 VIII

Bowden, J., Spiller, W., Del Greco M, F., Sheehan, N., Thompson, J., Minelli, C., Davey Smith, G. (2018). Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *International Journal of Epidemiology*, **47**, 1264–1278.

- 双向因果作用

Li, S., Ye, T. (2024). A focusing framework for testing bi-directional causal effects with GWAS summary data. *Journal of the Royal Statistical Society: Series B*.

Xie, F., Yao, Z., Xie, L., Zeng, Y., Geng, Z. (2024). Identification and estimation of the bi-directional MR with some invalid instruments. In *Proceedings of the 38th International Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada.

阴性对照变量

Negative Controls, Proximal Casual Inference

阴性对照下相关问题 I

$$
\begin{tikzcd}
\text{(NCE) } Z \arrow[dr] & & U \arrow[ll] \arrow[rr] \arrow[dl] \arrow[dr] & & W \text{ (NCO)} \arrow[dl] \\
& X \arrow[rr] & & Y &
\end{tikzcd}
$$

- 阴性对照识别

Kuroki, M., Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, **101**(2), 423–437.

Miao, W., Geng, Z., Tchetgen Tchetgen, E. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, **105**, 987–993.

Shi, X., Miao, W., Nelson, J. C., Tchetgen Tchetgen, E. (2020). Multiply robust causal inference with double negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B*, **82**, 521–540.

阴性对照下相关问题 II

- Miao, W., Shi, X., Li, Y., Tchetgen Tchetgen, E. J. (2024). A confounding bridge approach for double negative control inference on causal effects. *Statistical Theory and Related Fields*, **8**(4), 262–273.
- Cui, Y., Pu, H., Shi, X., Miao, W., Tchetgen Tchetgen, E. (2023). Semiparametric proximal causal inference. *Journal of the American Statistical Association*, **119**(546), 1348–1359.
- Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., Miao, W. (2024). An introduction to proximal causal inference. *Statistical Science*, **39**(3), 375–390.

- 阴性对照稳健估计

- Shi, X., Miao, W., Nelson, J. C., Tchetgen Tchetgen, E. (2020). Multiply robust causal inference with double negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B*, **82**, 521–540.

阴性对照下相关问题 III

Cui, Y., Pu, H., Shi, X., Miao, W., Tchetgen Tchetgen, E. (2023). Semiparametric proximal causal inference. *Journal of the American Statistical Association*, **119**(546), 1348–1359.

- 中介分析存在未知混杂的识别性

Dukes, O., Shpitser, I., Tchetgen Tchetgen, E. J. (2023). Proximal mediation analysis. *Biometrika*, **110**(4), 973–987.

- 溢出效应存在未知混杂的识别性

Egami, N., Tchetgen Tchetgen, E. J. (2024). Identification and estimation of causal peer effects using double negative controls for unmeasured network confounding. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **86**(2), 487–511.

阴性对照下相关问题 IV

- 死亡删失存在未知混杂的识别性

  Luo, S., Li, W., Miao, W., He, Y. (2024). Identification and estimation of causal effects in the presence of confounded principal strata. *Statistics in Medicine*, 2024; 43(22): 4372-4387.

- 数据融合/无效工具变量（NCE）

  Dukes, O., Richardson, D. B., Shahn, Z., Robins, J. M., Tchetgen Tchetgen, E. J. (2024). Using negative controls to identify causal effects with invalid instrumental variables. *Biometrika*, to appear.

- binary outcome (regression view) Proximal Causal Inference

阴性对照下相关问题 V

Liu, J., Park, C., Li, K., Tchetgen Tchetgen, E. J. (2024). Regression-based proximal causal inference. *American Journal of Epidemiology*, to appear.

- 混杂桥函数的非参估计

Li, W., Miao, W., Tchetgen Tchetgen, E. (2023). Nonparametric inference about mean functionals of nonignorable nonresponse data without identifying the joint distribution. *Journal of the Royal Statistical Society: Series B*, **85**(3), 913–935.

Zhang, J., Li, W., Miao, W., Tchetgen Tchetgen, E. (2023). Proximal causal inference without uniqueness assumptions. *Statistics & Probability Letters*, **198**, 109836.

阴性对照下相关问题 VI

Kallus, N., Mao, X., Uehara, M. (2021). Causal inference under unmeasured confounding with negative controls: A minimax learning approach. *arXiv preprint arXiv:2103.14029*.

Ghassami, A., Ying, A., Shpitser, I., Tchetgen Tchetgen, E. (2022). Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference. In *International Conference on Artificial Intelligence and Statistics* (pp. 7210–7239). PMLR.

- 阴性对照 OTR

Qi, Z., Miao, R., Zhang, X. (2023). Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, **119**(546), 915–928.

- 找阴性对照变量，因果发现

阴性对照下相关问题 VII

Shi, X., Miao, W., Tchetgen, E. T. (2020b). A selective review of negative control methods in epidemiology. *Current Epidemiology Reports*, **7**, 190–202.

Xie, F., Chen, Z., Luo, S., Miao, W., Cai, R., Geng, Z. (2024). Automating the selection of proxy variables of unmeasured confounders. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, Vienna, Austria.

Kummerfeld, E., Lim, J., Shi, X. (2024). Data-driven Automated Negative Control Estimation (DANCE): search for, validation of, and causal inference with negative controls. *Journal of Machine Learning Research*, **25**(229), 1–35.

- 选择偏差，Test-Negative Design

Schnitzer, M. E. (2022). Estimands and estimation of COVID-19 vaccine effectiveness under the test-negative design: Connections to causal inference. *Epidemiology (Cambridge, Mass.)*, **33**(3), 325–333.

阴性对照下相关问题 VIII

Li, K. Q., Shi, X., Miao, W., Tchetgen Tchetgen, E. (2024). Double negative control inference in test-negative design studies of vaccine effectiveness. *Journal of the American Statistical Association*, **119**(547), 1859–1870.

主分层分析

Principal Stratification Analysis

主分层分析下相关问题 I

- 基础知识

Ding, P. (2024). *A first course in causal inference*. CRC Press. (Chapter 26)

- 主分层因果作用的界

Zhang, J. L., Rubin, D. B. (2003). Estimation of causal effects via principle stratification when some outcomes are truncated by ‘death’. *Journal of Educational and Behavioral Statistics*, **28**, 353–368.

Imai, K. (2008). Sharp bounds on the causal effects in randomized experiments with “truncation-by-death”. *Statistics & Probability Letters*, **78**, 144–149.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, **76**, 1071–1102.

Long, D. M., Hudgens, M. G. (2013). Sharpening bounds on principal effects with covariates. *Biometrics*, **69**, 812–819.

主分层分析下相关问题 II

- 主分层因果作用的识别性

Ding, P., Geng, Z., Yan, W., Zhou, X.-H. (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association*, **106**, 1578–1591.

Zhang, J. L., Rubin, D. B., Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, **104**, 166–176.

Jiang, Z., Ding, P., Geng, Z. (2016). Principal causal effect identification and surrogate end point evaluation by multiple trials. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **78**, 829–848.

Ding, P., Lu, J. (2017). Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 757–777.

主分层分析下相关问题 III

- Wang, L., Richardson, T. S., Zhou, X.-H. (2017). Causal analysis of ordinal treatments and binary outcomes under truncation by death. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 719–735.
- Wang, L., Zhou, X.-H., Richardson, T. S. (2017). Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*, **104**, 597–612.
- Jiang, Z., Ding, P. (2021). Identification of causal effects within principal strata using auxiliary variables. *Statistical Science*, **36**, 493–508.
- Jiang, Z., Yang, S., Ding, P. (2022). Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **84**(4), 1423–1445.
- Luo, S., Li, W., He, Y. (2023). Causal inference with outcomes truncated by death in multiarm studies. *Biometrics*, **79**(1), 502–513.

主分层分析下相关问题 IV

Luo, S., Li, W., Miao, W., He, Y. (2024). Identification and estimation of causal effects in the presence of confounded principal strata. *Statistics in Medicine*, 2024; 43(22): 4372-4387.

- 归因分析

Pearl, J. (1999). Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, **121**(1), 93–149.

Shingaki, R., Kuroki, M. (2021). Identification and estimation of joint probabilities of potential outcomes in observational studies with covariate information. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 26475–26486.

Lu, Z., Geng, Z., Li, W., Zhu, S., Jia, J. (2023). Evaluating causes of effects by posterior effects of causes. *Biometrika*, **110**(2), 449–465.

主分层分析下相关问题 V

Kawakami, Y., Shingaki, R., Kuroki, M. (2023). Identification and estimation of the probabilities of potential outcome types using covariate information in studies with non-compliance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**(10), 12234–12242.

Li, W., Lu, Z., Jia, J., Xie, M., Geng, Z. (2024). Retrospective causal inference with multiple effect variables. *Biometrika*, **111**(2), 573–589.

- 公平性评价

Imai, K., Jiang, Z., Greiner, J., Halen, R., Shin, S. (2023). Experimental evaluation of algorithm-assisted human decision-making: Application to pretrial public safety assessment (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **186**, 167–189.

主分层分析下相关问题 VI

Imai, K., Jiang, Z. (2023). Principal fairness for human and algorithmic decision-making. *Statistical Science*, **38**, 317–328.

Ben-Michael, E., Greiner, D. J., Huang, M., Imai, K., Jiang, Z., Shin, S. (2024). Does AI help humans make better decisions? A methodological framework for experimental evaluation. *arXiv preprint*, arXiv:2403.12108.

- 敏感性分析

Schwartz, S., Li, F., Reiter, J. P. (2012). Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables. *Statistics in Medicine*, **31**(10), 949–962.

Ding, P., Lu, J. (2017). Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 757–777.

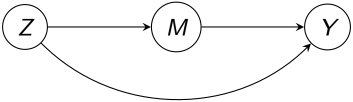
主分层分析下相关问题 VII

- Optimal treatment regime (OTR); Policy learning
  - Ben-Michael, E., Imai, K., Jiang, Z. (2024). Policy learning with asymmetric counterfactual utilities. *Journal of the American Statistical Association*, **2024**, 1–14.
- 与中介分析的联系
  - Forastiere, L., Mattei, A., Ding, P. (2018). Principal ignorability in mediation analysis: through and beyond sequential ignorability. *Biometrika*, **105**(4), 979–986.

中介分析

Mediation Analysis

中介分析下相关问题 I



- 基础知识
  - Ding, P. (2024). *A first course in causal inference*. CRC Press. (Chapters 27-28)
- 稳健估计
  - Tchetgen Tchetgen, E. J., Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of Statistics*, **40**(3), 1816.
- 多（高维）中介变量
  - VanderWeele, T., Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic Methods*, **2**, 95–115.

中介分析下相关问题 II

Daniel, R. M., De Stavola, B. L., Cousens, S., Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics*, **71**, 1–14.

Xia, F., Chan, K. C. G. (2022). Decomposition, identification and multiply robust estimation of natural mediation effects with multiple mediators. *Biometrika*, **109**(4), 1085–1100.

- 存在未知混杂的识别问题

Guo, Z., Small, D. S., Gansky, S. A., Cheng, J. (2018). Mediation analysis for count and zero-inflated count data without sequential ignorability and its application in dental studies. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **67**, 371–394.

中介分析下相关问题 III

- Yuan, Y., Qu, A. (2024). De-confounding causal inference using latent multiple-mediator pathways. *Journal of the American Statistical Association*, **119**(547), 2051–2065.
- Shuai, K., Liu, L., He, Y., Li, W. (2023). Mediation pathway selection with unmeasured mediator-outcome confounding. *arXiv preprint*, arXiv:2311.16793.
- Dukes, O., Shpitser, I., Tchetgen Tchetgen, E. J. (2023). Proximal mediation analysis. *Biometrika*, **110**(4), 973–987.

- 缺失数据

- Li, W., Zhou, X.-H. (2017). Identifiability and estimation of causal mediation effects with missing data. *Statistics in Medicine*, **36**, 3948–3965.
- Zuo, S., Ghosh, D., Ding, P., Yang, F. (2024). Mediation analysis with the mediator and outcome missing not at random. *Journal of the American Statistical Association*, to appear.

非随机缺失

Nonignorable Missing Data

缺失数据下相关问题 I

- 非随机缺失的识别，shadow variable

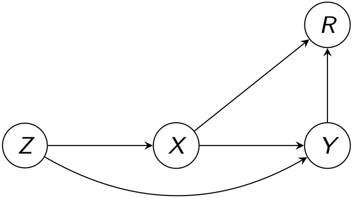


图: (a) $Z \perp\!\!\!\perp R \mid (Y, X)$; (b) $Z \not\perp\!\!\!\perp Y \mid (R = 1, X)$

Ma, W. Q., Geng, Z., Hu, Y. H. (2003). Identification of graphical models for nonignorable nonresponse of binary outcomes in longitudinal studies. *Journal of Multivariate Analysis*, **87**(1), 24–45.

缺失数据下相关问题 II

D' Haultfœuille, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics*, **154**, 1–15.

Wang, S., Shao, J., Kim, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, **24**, 1097–1116.

Ding, P., Geng, Z. (2014). Identifiability of subgroup causal effects in randomized experiments with nonignorable missing covariates. *Statistics in Medicine*, **33**(7), 1121–1133.

Miao, W., Tchetgen Tchetgen, E. J. (2016). On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika*, **103**, 475–482.

Li, W., Zhou, X.-H. (2017). Identifiability and estimation of causal mediation effects with missing data. *Statistics in Medicine*, **36**, 3948–3965.

缺失数据下相关问题 III

- Yang, S., Wang, L., Ding, P. (2019). Causal inference with confounders missing not at random. *Biometrika*, **106**(4), 875–888.
- Zhao, J., Ma, Y. (2022). A versatile estimation procedure without estimating the nonignorable missingness mechanism. *Journal of the American Statistical Association*, **117**(540), 1916–1930.
- Li, W., Miao, W., Tchetgen Tchetgen, E. (2023). Nonparametric inference about mean functionals of nonignorable nonresponse data without identifying the joint distribution. *Journal of the Royal Statistical Society: Series B*, **85**(3), 913–935.
- Zuo, S., Ghosh, D., Ding, P., Yang, F. (2024). Mediation analysis with the mediator and outcome missing not at random. *Journal of the American Statistical Association*, to appear.

- 非随机缺失的识别，instrumental variable

缺失数据下相关问题 IV

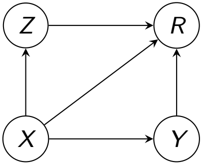


图: (a) $Z \not\perp\!\!\!\perp R \mid X$; (b) $Z \perp\!\!\!\perp Y \mid X$

Tchetgen Tchetgen, E. J., Wirth, K. E. (2017). A general instrumental variable framework for regression analysis with outcome missing not at random. *Biometrics*, **73**, 1123–1131.

Sun, B., Liu, L., Miao, W., Wirth, K., Robins, J., Tchetgen Tchetgen, E. J. (2018). Semiparametric estimation with data missing not at random using an instrumental variable. *Statistica Sinica*, **28**(4), 1965.

缺失数据下相关问题 V

Liu, L., Miao, W., Sun, B., Robins, J., Tchetgen Tchetgen, E. (2020). Identification and inference for marginal average treatment effect on the treated with an instrumental variable. *Statistica Sinica*, **30**(3), 1517.

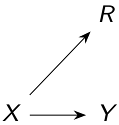
- 非随机缺失的识别，参数模型

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, **47**, 153–161.

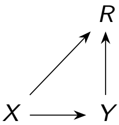
Miao, W., Ding, P., Geng, Z. (2016). Identifiability of normal and normal mixture models with nonignorable missing data. *Journal of the American Statistical Association*, **111**, 1673–1683.

随机缺失相关知识

缺失机制



MAR



MNAR

- 随机缺失 (Missing at random, MAR): $R \perp\!\!\!\perp Y \mid X$;
  (缺失只依赖于完全观测到的协变量，不依赖于缺失值本身)
- 非随机缺失 (Missing not at random, MNAR): $R \not\!\perp\!\!\!\perp Y \mid X$;
  (缺失是依赖于缺失值的)

随机缺失（MAR）

- 基于模型 $E(Y \mid X, R=1) = m(X;\beta)$ 的回归估计（REG）

$$\hat{\mu}_{\text{reg}} = \hat{E}\left\{m(X_i;\hat{\beta})\right\}$$

- 基于倾向评分模型 $\text{pr}(R=1 \mid X) = \pi(X;\alpha)$ 的逆概率加权估计（IPW）

$$\hat{\mu}_{\mathrm{ipw}} = \hat{E}\left\{\frac{R_i}{\pi\left(X_i;\hat{\alpha}\right)} Y_i\right\}$$

- 双稳健估计（AIPW）:

$$\begin{aligned}\hat{\mu}_{\text{aipw}} &= \hat{E}\left[\frac{R_i}{\pi(x;\hat{\alpha})} Y_i + \left\{1 - \frac{R}{\pi(X;\hat{\alpha})}\right\} m(X_i;\hat{\beta})\right] \\ &= \hat{E}\left[m(X_i;\hat{\beta}) + \frac{R_i}{\pi(x;\hat{\alpha})}\left\{Y - m(X_i;\hat{\beta})\right\}\right]\end{aligned}$$

未完待续...