PLSC 40601: Advanced Topics in Causal Inference.

University of Chicago, Spring 2023.

Location: Harper Memorial Library 104
Course time: Tue Thu 14.00–15:20

Instructor: Molly Offer-Westort; mollyow@uchicago.edu
Office hours: Book at calendly.com/mollyow
Office: Pick Hall 526

TA: Oscar Cuadros; oscarcuadros@uchicago.edu
Office hours: Monday/Wednesday 10:30 to 11:30 AM, at The Keller Center - Harris School of Public Policy. Please email the day before scheduling a meeting, with information on the topic you would like to discuss.

Course overview. This is a graduate-level course considering modern advances in causal inference and experimental design. In particular, we will consider how machine learning methods can be leveraged to address causal questions. We will read a selection of papers introducing and implementing techniques and research designs, with applications to the social and health sciences and public policy. We will discuss what these new methods are able to offer, and where they may have limitations. The course will be oriented around class discussion and student presentations on the readings. An introductory course in probability and statistics is required; this prerequisite can be met by courses in statistics, biostatistics, economics, political science, sociology, or related fields. Coursework in causal inference is recommended but not required; additional reading references will be provided for students who have not had prior exposure to causal inference methodology. Instructor consent required.

Reference readings. This class will not follow a textbook, but some of the methods and approaches may be new to students. For reference for machine learning methods, I recommend:

- Hastie, T., R. Tibshirani, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer

  A free pdf version of the book is available at the publisher's website: https://link.springer.com/book/10.1007/978-0-387-84858-7

Other machine learning references are :

- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York: Springer-Verlag

Many of the papers we read in this course will use potential outcomes notation. Students who have not had recent coursework on causal inference should read the first few chapters of either:

- Hernán, M. A. and J. M. Robins (2010). *Causal inference: What if?* CRC Boca Raton, FL
  A free pdf version of the book is available at the author's website: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
- Or
  Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press

As well the materials for PLSC 30600: Causal Inference are available at https://github.com/UChicago-pol-methods/plsc-30600-causal-inference.

Much of the statistical theory referenced in readings is covered in this book:

- Lehmann, E. L. (1999). *Elements of large-sample theory*. Springer

Also useful for econometrics references:

- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press

Presentations. Each student will be responsible for presenting and leading discussion on one of the required papers, beginning the second week of the quarter. Discussion leaders will provide a 15-20 minute total high-level overview of the paper, and will develop a list of discussion questions. For classes with multiple leaders, co-discussion leaders may divide papers among themselves or may collaborate. Discussion questions should be submitted to the instructor by email the evening before class.

Two pages referee report. Select one of the papers on the syllabus, including from the additional references. Write a referee report as if you were requested to review the paper for publication at a top journal, summarizing the paper's main contributions, areas that could be strengthened, and additional extensions or analyses that you would like to see. Submit the referee report on Canvas the night before the paper is discussed in class. The paper selected for the referee report cannot be from the same class as the class presentation. References on how to write a referee report:

- Berk, J., C. R. Harvey, and D. A. Hirshleifer (2016). Preparing a referee report: guidelines and perspectives. *Available at SSRN 2547191*
- https://chrisblattman.com/blog/2012/01/18/how-to-referee-an-academic-paper/

Five pages single spaced summary paper. (Seven pages hard cap, longer is not always better.) The paper may be summarizing work on one of the themes we have discussed, or on another topic related to developments in machine learning for causal inference. Discuss the foundations for this work in statistics or related fields, how this work has developed, competing methodologies or debates with associated benefits and drawbacks, and directions for future research. This should be something like an annotated bibliography, referencing each paper's contributions, accompanied by narrative. A reader of your report should come away with a good understanding of the current state of research on your topic of choice.

Alternatively, students may submit original work, such as a research proposal for a project they are working on. This option is available with instructor permission only. Paper topics and proposals for original work must be emailed to the instructor by the Wednesday of week 7 of the course. Submit the assignment by 11:59 PM on the Friday of finals week.

Grade composition is:

- Leading class discussions: 25%
- Referee report: 25%
- Final paper: 40%
- Participation: 10%

Accommodations: Please reach out to me directly if you would like to request accommodations for the course to better facilitate your learning. Student Disability Services (`https://disabilities.uchicago.edu/`) is also available to provide you resources and support, and may provide approval for specific academic accommodations. If you or your household is affected by the ongoing pandemic in a way that affects your ability to participate in or attend class, please reach out to me as well. Informing me in a timely manner will help me to ensure accommodations are met and I am able to implement an appropriate assessment of your learning.

Course outline.

Week 1. Causal inference, identification, course orientation.

Tuesday. Lecture.

- Murphy, J. R. (1997). How to read the statistical methods literature: a guide for students. *The American Statistician 51*(2), 155–157

Thursday. Lecture.

- Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science 16*(3), 199–231
  (comments and rejoinder optional)
- Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*, 685–725

Week 2. Covariate adjustment and balancing.

Tuesday.

- Bloniarz, A., H. Liu, C.-H. Zhang, J. S. Sekhon, and B. Yu (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences 113*(27), 7383–7390

- *Optional/reference:* Urminsky, O., C. Hansen, and V. Chernozhukov (2016). Using double-lasso regression for principled variable selection. *Available at SSRN 2733374*
- *Optional/reference:* Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650
- *Optional/reference:* Athey, S., G. W. Imbens, and S. Wager (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(4), 597–623

Thursday. Lecture: sample-splitting, bagging, honesty.

- *Optional/reference:* Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological) 36*(2), 111–133
- *Optional/reference:* Breiman, L. (1996). Bagging predictors. *Machine learning 24*, 123–140
- *Optional/reference:* Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist. 7*(1), 1–26

Week 3. Heterogeneous treatment effects.

Tuesday. Lecture: Trees and forests.

- Breiman, L. (2001a). Random forests. *Machine learning 45*, 5–32

Thursday.

- Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242
- *Optional/reference:* Athey, S. and G. Imbens (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*(27), 7353–7360
- *Optional/reference:* Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests

Week 4. More heterogeneous treatment effects, policy learning.

Tuesday.

- Künzel, S. R., J. S. Sekhon, P. J. Bickel, and B. Yu (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences 116*(10), 4156–4165

Thursday.

- EITHER
  Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*(1), 133–161
- OR
  Kitagawa, T. and A. Tetenov (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica 86*(2), 591–616
  (you don't need to read both)

Week 5. Efficient nonparametric parameter estimation.

Tuesday.

- Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469.
  Parts 1-3.

- *Optional/reference:* Edward Kennedy's slides from the ACIC conference in 2022: *Machine learning & nonparametric efficiency in causal inference* https://www.ehkennedy.com/uploads/5/8/4/5/58450265/tutorial.pdf
  The slides present the same material as the paper, with some additional narrative & figures, and references throughout to relevant historical work. There are also some videos of Kennedy giving lectures on this topic on YouTube, e.g., https://youtu.be/vZKdObmilQU.
- *Optional/reference:* Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the american statistical association 69*(346), 383–393
- *Optional/reference:* Lehmann, E. L. (1999). *Elements of large-sample theory*. Springer
  In particular, sections 1.1-1.4 on limits and rates of convergence. Fisher Information is covered in sections 7.1 and 7.2. The theory of statistical functionals and the influence curve is covered in sections 6.2 and 6.3.

Thursday.

- Kennedy (2022) continued.
  Parts 4-5.

Week 6. Double Machine learning.

Tuesday.

- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018b). Double/debiased machine learning for treatment and structural

parameters: Double/debiased machine learning. *The Econometrics Journal 21*(1)
Parts 1-3.

- *Optional/reference:* Victor Chernozhukov's slides from a 2016 talk at UChicago: *Double Machine Learning for Causal and Treatment Effects*, https://bfi.uchicago.edu/wp-content/uploads/4A_Victor_talk_DoubleML.pdf
- *Optional/reference:* Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey (2017). Double/Debiased/Neyman Machine Learning of Treatment Effects. *American Economic Review 107*(5), 261–65
- *Optional/reference:* Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018a, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68

Thursday.

- Chernozhukov et al. (2018b) continued.
Parts 4-6.

Week 7. Targeted maximum likelihood estimation.

Tuesday.

- Schuler, M. S. and S. Rose (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology 185*(1), 65–73

- *Optional/reference:* Katherine Hoffman's *An Illustrated Guide to TMLE*, https://www.khstats.com/blog/tmle/tutorial
- *Optional/reference:* Gruber, S. and M. J. Van Der Laan (2009). Targeted maximum likelihood estimation: A gentle introduction
- *Optional/reference:* Van Der Laan, M. J. and D. Rubin (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics 2*(1)
- *Optional/reference:* Gruber, S. and M. Van Der Laan (2012). tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software 51*, 1–35

Thursday.

- Review of recent methods.

- *Optional/reference:* Mark van der Laan's lab blog, *CV-TMLE and double machine learning*, `https://vanderlaan-lab.org/2019/12/24/cv-tmle-and-double-machine-learning`
- *Optional/reference:* Díaz, I. (2020). Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics 21*(2), 353–358

Week 8. Reinforcement learning and adaptive experiments.

Tuesday.

- Bubeck, S., R. Munos, and G. Stoltz (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science 412*(19), 1832–1852

- *Optional/reference:* Russo, D., B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen (2017, July). A Tutorial on Thompson Sampling. arXiv e-prints arXiv:1707.02038
- *Optional/reference:* Lai, T. L. and H. Robbins (1985). Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics 6*(1), 4–22

Thursday.

- Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences 118*(15)

- *Optional/reference:* Zhan, R., V. Hadad, D. A. Hirshberg, and S. Athey (2021). Off-policy evaluation via adaptive weighting with data from contextual bandits. *arXiv preprint arXiv:2106.02029*

Week 9. Conformal inference

Tuesday.

- Lei, L. and E. J. Candès (2021). Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology 83*(5), 911–938
- Cyrus Samii's conformal tutorial: `https://cdsamii.github.io/cds-demos/conformal/conformal-tutorial.html`

- *Optional/reference:* Vovk, V., A. Gammerman, and G. Shafer (2005). *Algorithmic learning in a random world*. Springer Science & Business Media
- *Optional/reference:* Lei, J., M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association 113*(523), 1094–1111
- *Optional/reference:* Chernozhukov, V., K. Wüthrich, and Y. Zhu (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association 116*(536), 1849–1864

Thursday.

- Open/discussion.

References

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Athey, S. and G. Imbens (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*(27), 7353–7360.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*, 685–725.

Athey, S., G. W. Imbens, and S. Wager (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(4), 597–623.

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests.

Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*(1), 133–161.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.

Berk, J., C. R. Harvey, and D. A. Hirshleifer (2016). Preparing a referee report: guidelines and perspectives. *Available at SSRN 2547191*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York: Springer-Verlag.

Bloniarz, A., H. Liu, C.-H. Zhang, J. S. Sekhon, and B. Yu (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences 113*(27), 7383–7390.

Breiman, L. (1996). Bagging predictors. *Machine learning 24*, 123–140.

Breiman, L. (2001a). Random forests. *Machine learning 45*, 5–32.

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science 16*(3), 199–231.

Bubeck, S., R. Munos, and G. Stoltz (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science 412*(19), 1832–1852.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey (2017). Double/Debiased/Neyman Machine Learning of Treatment Effects. *American Economic Review 107*(5), 261–65.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018a, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018b). Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal 21*(1).

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association 116*(536), 1849–1864.

Díaz, I. (2020). Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics 21*(2), 353–358.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist. 7*(1), 1–26.

Gruber, S. and M. Van Der Laan (2012). tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software 51*, 1–35.

Gruber, S. and M. J. Van Der Laan (2009). Targeted maximum likelihood estimation: A gentle introduction.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences 118*(15).

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the american statistical association 69*(346), 383–393.

Harshaw, C., F. Sävje, D. Spielman, and P. Zhang (2019). Balancing covariates in randomized experiments with the gram–schmidt walk design. *arXiv preprint arXiv:1911.03071*.

Hastie, T., R. Tibshirani, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.

Hernán, M. A. and J. M. Robins (2010). *Causal inference: What if?* CRC Boca Raton, FL.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469.

Kitagawa, T. and A. Tetenov (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica 86*(2), 591–616.

Künzel, S. R., J. S. Sekhon, P. J. Bickel, and B. Yu (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences 116*(10), 4156–4165.

Lai, T. L. and H. Robbins (1985). Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics 6*(1), 4–22.

Lehmann, E. L. (1999). *Elements of large-sample theory*. Springer.

Lei, J., M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association 113*(523), 1094–1111.

Lei, L. and E. J. Candès (2021). Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology 83*(5), 911–938.

Murphy, J. R. (1997). How to read the statistical methods literature: a guide for students. *The American Statistician 51*(2), 155–157.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Russo, D., B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen (2017, July). A Tutorial on Thompson Sampling. arXiv e-prints arXiv:1707.02038.

Schuler, M. S. and S. Rose (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology 185*(1), 65–73.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological) 36*(2), 111–133.

Urminsky, O., C. Hansen, and V. Chernozhukov (2016). Using double-lasso regression for principled variable selection. *Available at SSRN 2733374*.

Van Der Laan, M. J. and D. Rubin (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics 2*(1).

Vovk, V., A. Gammerman, and G. Shafer (2005). *Algorithmic learning in a random world*. Springer Science & Business Media.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

Xu, Y. and E. Yang (2022). Hierarchically regularized entropy balancing. *Political Analysis*, 1–8.

Zhan, R., V. Hadad, D. A. Hirshberg, and S. Athey (2021). Off-policy evaluation via adaptive weighting with data from contextual bandits. *arXiv preprint arXiv:2106.02029*.