

STAT 9700: TOPICS FOR FINAL PROJECT

1. *Estimation in sampled and noisy networks:*

- J. M. Klusowski and Y. Wu, Counting motifs with graph sampling, *Conference on Learning Theory (COLT)*, 2018.
- J. M. Klusowski and Y. Wu, Estimating the number of connected components in a graph via subgraph sampling, *Bernoulli*, Vol. 26 (3), 1635–1664, 2020.
- B. B. Bhattacharya, S. Das, and S. Mukherjee, Motif estimation via subgraph sampling: the fourth moment phenomenon, *Annals of Statistics*, Vol. 50, 987–101, 2022.

2. *Graphon Estimation:*

- Sourav Chatterjee, Matrix estimation by universal singular value thresholding, *Annals of Statistics*, Vol. 43 (1), 177–214, 2015.
- C. Gao, Y. Lu, and H. Zhou, Rate-optimal graphon estimation, *Annals of Statistics*, Vol. 43 (6), 2624–2652, 2015.
- O. Klopp, A. B. Tsybakov, and N. Verzelen, Oracle inequalities for network models and sparse graphon estimation, *Annals of Statistics*, Vol. 45 (1), 316–354, 2017.

3. *Conditional Dependence Measures:*

- M. Azadkia and S. Chatterjee, A simple measure of conditional dependence, *Annals of Statistics*, Vol. 49 (6), 3070–3102, 2021.
- Z. Huang, N. Deb, and B. Sen, Kernel partial correlation coefficient – a measure of conditional dependence, *Journal of Machine Learning Research*, Vol. 23, 1–58, 2022.

4. *Estimation in the β -model:*

- S. Chatterjee, P. Diaconis, and A. Sly, Random graphs with a given degree sequence, *Annals of Applied Probability*, Vol. 21, 1400–1435, 2011.
- A. Rinaldo, S. Petrović, and S. E. Fienberg, Maximum likelihood estimation in the β -model, *Annals of Statistics*, Vol. 41, 1085–1110, 2013.
- M. Chen, K. Kato, and C. Leng, Analysis of networks via the sparse β -model, *Journal of the Royal Statistical Society, Series B*, Vol. 83 (5), 887–910, 2021.

5. *Sequential Nonparametric Testing:*

- S. Shekhar and A. Ramdas, Nonparametric two-sample testing by betting, *IEEE Transactions on Information Theory*, [arXiv:2112.09162](https://arxiv.org/abs/2112.09162), 2023.
- A. Podkopaev, P. Blöbaum, S. Kasiviswanathan, and A. Ramdas, Sequential kernelized independence testing, *International Conference on Machine Learning*, 27957–27993, 2023.

6. *Conformal Inference:*

- J. Lei, J. Robins, and L. Wasserman, Distribution-free prediction sets, *Journal of the American Statistical Association*, Vol. 108 (501), 278–287, 2013.
- R. F. Barber E. J. Candès A. Ramdas R. J. Tibshirani, Predictive inference with the jackknife+, *Annals of Statistics*, Vol. 49(1), 486–507, 2021.
- A. N. Angelopoulos and S. Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification, [arXiv:2107.07511](#), 2022.
- R. F. Barber E. J. Candès A. Ramdas R. J. Tibshirani, Conformal prediction beyond exchangeability, *Annals of Statistics*, 816–845, 2023.

7. *Network Method of Moments:*

- P. Bickel, A. Chen, and E. Levina, The method of moments and degree distributions for network models, *Annals of Statistics*, Vol. 39, 2280–2301, 2011.
- Y. Zhang and D. Xia, Edgeworth expansions for network moments, *Annals of Statistics*, Vol. 50, 726–753, 2022.

8. *High-Dimensional Variational and Empirical Bayes Estimation:*

- S. Mukherjee and S. Sen, Variational inference in high-dimensional linear regression, *Journal of Machine Learning Research*, Vol. 23, 1–56, 2022.
- Y. Kim, W. Wang, P. Carbonetto, and M. Stephens, A flexible empirical Bayes approach to multiple linear regression and connections with penalized regression, [arXiv:2208.10910](#), 2022.
- S. Mukherjee, B. Sen, and S. Sen, A mean field approach to empirical Bayes estimation in high-dimensional linear regression, [arXiv:2309.16843](#), 2023.

9. *Sampling Subgraphs and Learning Network Structures:*

- H. Lyu, F. Memoli, and D. Sivakoff, Sampling random graph homomorphisms and applications to network data analysis, *Journal of Machine Learning Research*, Vol. 24, 1–79, 2023.
- H. Lyu, Y. Kureh, J. Vendrow, and M. A. Porter, Learning low-rank mesoscale structures of networks, *Nature Communications*, [arXiv:2102.06984](#), 2023.

10. *Mixture Models:*

- A. Kalai, A. Moitra, and G. Valiant, Disentangling gaussians, *Communications of the ACM*, Vol. 55, 113–120, 2012.
- Y. Wu and P. Yang, Optimal estimation of Gaussian mixtures via denoised method of moments, *Annals of Statistics*, Vol. 48, 1981–2007, 2020.
- C. Mao and Y. Wu, Learning mixtures of permutations: Groups of pairwise comparisons and combinatorial method of moment, *Annals of Statistics*, Vol. 50 (4), 2231–2255, 2022.

11. *Distributional Random Forests:*

- D. Cévid, L. Michel, J. Näf, P. Bühlmann, and N. Meinshausen, Distributional random forests: Heterogeneity adjustment and multivariate distributional regression, *Journal of Machine Learning Research*, Vol. 23, 1–79, 2022.
- J. Näf, C. Emmenegger, P. Bühlmann, and N. Meinshausen, Confidence and uncertainty assessment for distributional random forests, [arXiv:2302.05761](#), 2023.

12. *Sensitivity Analysis:*

- Q. Zhao, D. S. Small, and B. B. Bhattacharya, Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap, *Journal of the Royal Statistical Society: Series B*, Vol. 81, 735–761, 2019.
- J. Dorn and K. Guo, Sharp sensitivity analysis for inverse propensity weighting via quantile balancing, *Journal of the American Statistical Association*, 1–13, 2022.

13. *Neural Networks for Estimation:*

- B. Bauer and M. Kohler, On deep learning as a remedy for the curse of dimensionality in nonparametric regression, *Annals of Statistics*, Vol. 47 (4), 2261–2285, 2019.
- M. Farrell, T. Liang, and S. Misra, Deep Neural Networks for Estimation and Inference, *Econometrica*, Vol. 89, 181–213, 2021.

14. *Classifier Two-Sample Tests:*

- David Lopez-Paz and Maxime Oquab, Revisiting classifier two-sample tests, *International Conference on Learning Representations (ICLR)*, 2017.
- I. Kim, A. Ramdas, A. Singh, and L. Wasserman, Classification accuracy as a proxy for two-sample testing, *Annals of Statistics*, Vol. 49, 411–434, 2021.

15. *Conditional Independence Testing:*

- R. D. Shah and J. Peters, The hardness of conditional independence testing and the generalised covariance measure, *Annals of Statistics*, Vol. 48(3), 1514–1538, 2020.
- I. Kim, M. Neykov, S. Balakrishnan, and L. Wasserman, Local permutation tests for conditional independence, *Annals of Statistics*, Vol. 50 (6), 3388–3414, 2022.

16. *Confidence Interval for High-Dimensional Regression:*

- A. Javanmard and A. Montanari, Confidence intervals and hypothesis testing for high-dimensional regression, *The Journal of Machine Learning Research*, Vol. 15 (1), 2869–2909, 2014.
- T. T. Cai and Z. Guo, Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity, *Annals of Statistics*, Vol. 45 (2), 615–646, 2017.
- T. T. Cai, Z. Guo, and R. Ma, Statistical inference for high-dimensional generalized linear models with binary outcomes, *Journal of the American Statistical Association*, Vol. 118, 1319–1332, 2023.

17. *Thresholds in Sparse Linear Regression and Random Graphs:*

- D. Gamarnik and I. Zadik, Sparse high-dimensional linear regression. Estimating squared error and a phase transition, *Annals of Statistics*, Vol. 50 (2), 880–903, 2022.
- E. Mossel, J. Niles-Weed, Y. Sohn, N. Sun, and I. Zadik, Sharp thresholds in inference of planted subgraphs, [arXiv:2302.14830](#), 2023.

18. *Data Visualization and Aggregation Methods:*

- E. Sun, R. Ma, and J. Zou, Dynamic Visualization of High-Dimensional Data, *Nature Computational Science*, Vol. 3, 86–100, 2023.
- R. Ma, E. Sun, and J. Zou, A spectral method for assessing and combining multiple data visualizations, *Nature Communications*, 2023.

19. *Spectrum Estimation:*

- W. Kong and G. Valiant, Spectrum estimation from samples, *Annals of Statistics*, Vol. 45, No. 5, 2218–2247, 2017.
- Y. Jin, C. Musco, A. Sidford, A. V. Singh, Moments, random walks, and limits for spectrum approximation, *Proceedings of Thirty Sixth Conference on Learning Theory*, PMLR 195, 5373–5394, 2023.

20. *Graph Matching:*

- Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations I: The Gaussian model, *Foundations of Computational Mathematics*, to appear, 2022.
- Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality, *Foundations of Computational Mathematics*, to appear, 2023.
- C. Mao, Y. Wu, J. Xu, and S. H. Yu, Testing network correlation efficiently via counting trees, *Annals of Statistics*, to appear, 2023.

21. *Double Descent Phenomenon:*

- M. Belkin, D. Hsu, S. Ma, and S. Mandal, Reconciling modern machine learning and the bias-variance trade-off, *Proceedings of the National Academy of Sciences*, Vol. 116, 15849–15854, 2019.
- T. Hastie, A. Montanari, S. Rosset, and R. J. Tibshirani, Surprises in high-dimensional ridgeless least squares interpolation, Vol. 50, 949–986, 2022.