Statistical Inference for High-Dimensional Differential Networks

Byol Kim[1], Song Liu[2], and Mladen Kolar[1]
[1]University of Chicago, [2]University of Bristol

We study a graphical version of a two-sample statistical inference problem. Specifically, given a pair of independent samples, $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n_x)}$ and $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(n_y)}$, each from a member of a fixed parametric family of pairwise Markov random fields, we are interested in developing a statistically valid procedure for testing whether the $\boldsymbol{x}$-graph and the $\boldsymbol{y}$-graph differ for general pairwise Markov random fields that are not necessarily Gaussian and even when the possible number of edges p far outnumber the sample sizes n_x or n_y. To this end, we propose a method for constructing an asymptotically normal and unbiased estimator of the change in a scalar edge parameter θ_1 of interest.

Our proposed method builds on both the density ratio estimation literature and recent developments in post-selection or post-regularization inference. In particular, we make use of the sparse KLIEP procedure proposed in [1] to obtain an initial direct estimate of the change, to which we apply one-step correction, as suggested in [2], or a method that generalizes double selection, which was proposed in [3]. In simulation studies, we are able to demonstrate that the resulting estimator compares favorably with the oracle estimator.

Table 1: Simulation Results

(a) Empirical Coverage of Normal-Approximation-Based CIs

graph pair	m	n_x	n_y	oracle	naïve	one-step	double
chain #1	25	150	300	0.952	0.882	0.952	0.952
	50	300	600	0.955	0.836	0.959	0.950
chain #2	25	150	300	0.944	0.888	0.934	0.942
	50	300	600	0.936	0.817	0.920	0.939
3-ary tree #1	25	150	300	0.940	0.893	0.956	0.941
	50	300	600	0.945	0.865	0.963	0.956
3-ary tree #2	25	150	300	0.951	0.919	0.963	0.941
	50	300	600	0.942	0.858	0.951	0.936

(b) Empirical Bias $\times 10^2$

graph pair	m	n_x	n_y	oracle	naïve	one-step	double
chain #1	25	150	300	-0.396	7.011	-1.585	-1.428
	50	300	600	0.096	6.979	-1.469	-1.388
chain #2	25	150	300	-0.452	6.662	-1.144	-1.035
	50	300	600	-0.330	7.000	-0.814	-0.614
3-ary tree #1	25	150	300	-0.118	7.709	-0.968	-2.348
	50	300	600	-0.936	7.368	-0.025	-1.197
3-ary tree #2	25	150	300	0.252	6.455	0.793	0.225
	50	300	600	-0.487	8.234	-0.458	-0.858

References

[1] Song Liu, Taiji Suzuki, Raissa Relator, Jun Sese, Masashi Sugiyama, and Kenji Fukumizu, "Support consistency of direct sparse-change learning in markov networks," *The Annals of Statistics*, vol. 45, no. 3, pp. 959–990, jun 2017.

[2] Yang Ning and Han Liu, "A general theory of hypothesis tests and confidence regions for sparse high dimensional models," *The Annals of Statistics*, vol. 45, no. 1, pp. 158–195, feb 2017.

[3] Alexandre Belloni, Victor Chernozhukov, and Ying Wei, "Post-selection inference for generalized linear models with many controls," *Journal of Business & Economic Statistics*, vol. 34, no. 4, pp. 606–619, sep 2016.