Statistical Connectomics

jovo* cep†

August 29, 2014

Contents

Abstract

*yummy
†t

1 Introduction

additional core authors: runze?

potential additional graphstat authors: li? nam? youngser? daniele, dunson, vince, minh, daniel

potential neuro co-authors could include: mike milham, scott cook, mitya, bobby/jeff, clay/davi, rex jung,

we start by stating how important connectomics will be for the future of neuroscience, and how having rigorous statistical theory will enable future investigations to leverage it to substantiate their claims.

for each exploitation task, we provide:

1. rigorous definition
2. motivating application
3. R code
4. images, graphs, and graph derivatives downloads

2 One Sample Tests

2.1 tests for independence between connectivity and vertex attributes (such as direction preference, excitatory vs. inhibitory, etc.)

bock11 [1] dataset, testing independence of tuning direction vs connectivity, using residual error of regression o ase as test statistic, permutation test to obtain null

2.2 tests for independence between space and connectivity

kasthuri11 dataset (no cite yet, coming soon), touches vs. synapses, using whatever we do (probably importance sampling to obtain null distribution)

2.3 tests for model fit

hsbm on fly optic lobe data [2], likelihood test via parametric bootstrap

3 2-sample tests for comparative connectomics

3.1 comparing 2 different connectomes

elegans electrical vs. chemical & elegans vs. pacificus & elegans male vs. herm. See [3] for the most clear description of these graphs.

3.2 2 populations of connectomes

[4, 5] describes two different populations of subjects collected for two different studies, both of which are useful.

4 population density estimation

4.1 mean estimation

[6, **?**] are two papers proving that Stein's paradox does not occur in finite spaces, in other words, $\bar{A}$ is admissible under squared error loss. nonetheless, it seems likely that some smoothing/regularizing of $\bar{A}$ would be advantageous for finite sample sizes. in particular, spectral and constrained estimates of latent vectors. we can use any number of MR datasets, such as those MRN-111 in [7].

4.2 robust mean (eg, median, or Lq) estimation

we can again use the MRN-111 dataset, the theory is motivated by [8, 9].

4.3 Clustering

using tensor factorizations [10, 11, 12], or DELTACON [13, 14], which is just hclust with a different dissimilarity function.

4.4 errorbars around mean estimation, eg, estimation variance

bayesian nonparametric model [**?**]

5 connectome coding

5.1 classifying connectomes

signal subgraphs paper [15], or using ASE or tensor factorization, followed by classical classification.

5.2 regressing connectomes

MRN114 via NTF followed by regression onto CCI

5.3 multivariate regression for connectomes

Adelstein [16] using JoFC on 5-factor personality test.

6 Discussion

general issues:

6.1 bias variance trade-off: num params > num subjects

6.2 nuisance signals: age, sex, batch

6.3 Graph Matching

oh, which papers to list, how about [17, 18, 19, 20]

6.4 Future Work

Acknowledgments

The author thanks the anonymous authors whose work largely constitutes this sample file. He also thanks the INFO-TeX mailing list for the valuable indirect assistance he received.

References

[1] D. D. Bock, W.-C. A. Lee, A. M. Kerlin, M. Andermann, A. W. Wetzel, S. Yurgenson, E. R. Soucy, H. S. Kim, G. Hood, and R. C. Reid, "Network anatomy and in vivo physiology of visual cortical neurons," *Nature*, vol. 471, no. 7337, pp. 177–182, Mar. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3095821&tool=pmcentrez&rendertype=abstract

[2] S.-Y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, T. Zhao, J. A. Horne, R. D. Fetter, S. Takemura, K. Blazek, L.-A. Chang, O. Ogundeyi, M. A. Saunders, V. Shapiro, C. Sigmund, G. M. Rubin, L. K. Scheffer, I. A. Meinertzhagen, and D. B. Chklovskii, "A visual motion detection circuit suggested by Drosophila connectomics," *Nature*, vol. 500, no. 7461, pp. 175–181, Aug. 2013. [Online]. Available: http://dx.doi.org/10.1038/nature12450

[3] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, D. B. Chklovskii, C. Spring, and J. Farm, "Structural Properties of the Caenorhabditis elegans Neuronal Network," *PLoS Computational Biology*, vol. 7, no. 2, pp. 1–41, Feb. 2011.

[4] K. B. Nooner, S. J. Colcombe, R. H. Tobe, M. Mennes, M. M. Benedict, A. L. Moreno, L. J. Panek, S. Brown, S. T. Zavitz, Q. Li, S. Sikka, D. Gutman, S. Bangaru, R. T. Schlachter, S. M. Kamiel, A. R. Anwar, C. M. Hinz, M. S. Kaplan, A. B. Rachlin, S. Adelsberg, B. Cheung, R. Khanuja, C. Yan, R. C. Craddock, V. D. Calhoun, W. Courtney, M. King, D. Wood, C. L. Cox, A. M. C. Kelly, A. Di Martino, E. Petkova, P. T. Reiss, N. Duan, D. Thomsen, B. Biswal, B. Coffey, M. J. Hoptman, D. C. Javitt, N. Pomara, J. J. Sidtis, H. S. Koplewicz, X. F. Castellanos, B. L. Leventhal, and M. P. Milham, "The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry." *Frontiers in neuroscience*, vol. 6, p. 152, Jan. 2012. [Online]. Available: http://www.frontiersin.org/Brain_Imaging_Methods/10.3389/fnins.2012.00152/full

[5] B. a. Landman, A. J. Huang, A. Gifford, D. S. Vikram, I. A. L. Lim, J. a. D. Farrell, J. a. Bogovic, J. Hua, M. Chen, S. Jarso, S. a. Smith, S. Joel, S. Mori, J. J. Pekar, P. B. Barker, J. L. Prince, and P. C. M. van Zijl, "Multi-Parametric Neuroimaging Reproducibility: A 3T Resource Study." *NeuroImage*, Nov. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21094686

[6] B. M. Johnson, "On the Admissible Estimators for Certain Fixed Sample Binomial Problems," *The Annals of Mathematical Statistics*, vol. 42, no. 5, pp. 1579–1587, Oct. 1971. [Online]. Available: http://projecteuclid.org/euclid.aoms/1177693156

[7] W. Gray Roncal, Z. H. Koterba, D. Mhembere, D. M. Kleissas, J. T. Vogelstein, R. Burns, A. R. Bowles, D. K. Donavos, S. Ryman, R. E. Jung, L. Wu, V. D. Calhoun, and R. J. Vogelstein, "MIGRAINE: MRI Graph Reliability Analysis and Inference for Connectomics," *Global Conference on Signal and Information Processing*, 2013. [Online]. Available: http://arxiv.org/abs/1312.4875

[8] D. Ferrari and Y. Yang, "Maximum L q -likelihood estimation," *The Annals of Statistics*, vol. 38, no. 2, pp. 753–783, Apr. 2010. [Online]. Available: http://projecteuclid.org/euclid.aos/1266586613

[9] Y. Qin and C. E. Priebe, "Maximum L q -Likelihood Estimation via the Expectation-Maximization Algorithm: A Robust Estimation of Mixture Models," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 914–928, Sep. 2013. [Online]. Available: http://www.tandfonline.com.proxy3.library.jhu.edu/doi/abs/10.1080/01621459.2013.787933#.UuJtQmQo7og

[10] N. H. Lee, C. E. Priebe, R. Tang, and M. Rosen, "Using non-negative factorization of time series of graphs for learning from an event-actor network," Dec. 2013. [Online]. Available: http://arxiv.org/abs/1312.7559

[11] N. H. Lee, I.-J. Wang, Y. Park, C. E. Priebe, and M. Rosen, "Automatic Dimension Selection for a Non-negative Factorization Approach to Clustering Multiple Random Graphs," Jun. 2014. [Online]. Available: http://arxiv.org/abs/1406.6315

[12] N. H. Lee, I.-J. Wang, R. Tang, M. Rosen, and C. E. Priebe, "A rank estimation criterion using an NMF algorithm under an inner dimension condition," Jun. 2014. [Online]. Available: http://arxiv.org/abs/1406.6319

[13] D. Koutra, J. T. Vogelstein, and C. Faloutsos, "DeltaCon: Measuring Connectivity Differences in Large Networks," in *SIAM International Conference on Data Mining*, 2013. [Online]. Available: http://knowledgecenter.siam.org/105SDM/1

[14] D. Koutra, N. Shah, J. T. Vogelstein, B. J. Gallagher, and C. Faloutsos, "DELTACON: A Principled Massive-Graph Similarity Function and Applications," *in preparation.*

[15] J. T. Vogelstein, W. Gray Roncal, R. J. Vogelstein, and C. E. Priebe, "Graph Classification using Signal Subgraphs: Applications in Statistical Connectomics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1539 – 1551, 2013. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6341752

[16] J. S. Adelstein, Z. Shehzad, M. Mennes, C. G. DeYoung, X.-N. Zuo, C. Kelly, D. S. Margulies, A. Bloomfield, J. R. Gray, F. X. Castellanos, and M. P. Milham, "Personality is reflected in the brain's intrinsic functional architecture," *PLoS ONE*, vol. 6, no. 11, 2011. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-82455217247&partnerID=40&md5=15a8e206a9ebdcbe0e137dd8a34cc332

[17] J. T. Vogelstein, J. C. M. Conroy, L. J. Podrazik, S. G. Kratzer, D. E. Fishkind, R. J. Vogelstein, and C. E. Priebe, "(Brain) Graph Matching via Fast Approximate Quadratic Programming," *Submitted to Computational Statistics and Data Analysis*, 2013. [Online]. Available: http://arxiv.org/abs/1112.5507

[18] D. E. Fishkind, S. Adali, and C. E. Priebe, "Seeded Graph Matching," *arXiv preprint*, p. 1209.0367v1, 2012. [Online]. Available: http://arxiv.org/abs/1209.0367

[19] V. Lyzinski, S. Adali, J. T. Vogelstein, Y. Park, and C. E. Priebe, "Seeded Graph Matching Via Joint Optimization of Fidelity and Commensurability," *Submitted to Journal of Classification.*

[20] V. Lyzinski, D. L. Sussman, D. E. Fishkind, H. Pao, and C. E. Priebe, "Seeded graph matching for large stochastic block model graphs," *arXiv preprint*, p. 1310.1297, Oct. 2013. [Online]. Available: http://arxiv.org/abs/1310.1297