Tony Cai is good at statistics , and the statistical inferences in high-dimension may be useful

【8.11】

- Main references

[1] Donoho,D.L.& Jin.J.(2004).Higher criticisms for detecting, sparse and heterogeneous mixtures. Ann. Statist

Detection of homoscedastic Gaussian mixtures

[2] Cai,T.T., Jeng, J. & Jin. J. (2011) Optimal detection of heterogeneous and heteroscedastic mixtures J. Roy. Statist. Soc

Detection of heterogeneous & heteroscedastic Gaussian mixtures

[3] Cai. T.T & Sun. W. (2017). Optimal screening and discovery of sparse signals with applications to multistage high-throughput studies. J.Roy. Statist. Soc. Ser. B.79.197-233

Screening & discovery of sparse signals

[4] Jeng. J. Cai. T.T. &Li, H.(2010) Optimal sparse segment identification with application in copy number variation analysis. J. Amer. Statist. Assoc 105.1156-1166

Detection of segments-applications to CNV analysis

[5] Butucea,C. & Ingster, Yu. I. (2013) Detection of a sparse submatrix of a high-dimensional noisy matrix. Bernoulli 19. 2652-2688

Detection of submatrices

[6] Cai, T.T. &Wu. Y. (2014) Optimal detection for sparse mixture against a given null distribution. IEEE Trans. Inf. Theory 60

Detection of general (not necessarily Gaussian) mixtures

[7] Hall, P & Jin. J. (2008) Properties of Higher Criticism under long range dependence. Ann. Statist. 36

Detection of segments based on multiple sequences

[8] Hall, P & Jin. J. (2010) Innovated Higher Criticism for detecting sparse signals in correlated noise. Ann. Statist. 38

[9] Jeng. J. , Cai, T.T., Li. H .(2013) Simultaneous discovery of rare and common segment variants. Biometrika 100

[8]-[9] Detection under dependency

[10] Cai, T.T., Jin. J. & Low, M.G.(2007) Estimation and confidence sets for sparse normal mixtures. Ann. Statist. 35

[11] Jin. J. & Cai, T.T.(2007) Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. J. Amer. Statist. Assoc. 102

[12] Cai, T.T. & Jin, J. (2010) Optimal rates of convergences for estimating the null and proportion of non-null effects in large-scale multiple testing. Ann. Statist . 38

[10]-[12] Estimation of proportion (and null distribution)

- Theme

Talking about big data and big values, showed many applications of big data, especially about biology

【8.12】

- Theme
    1. Questions: how can we inferred whether a distribution is detectable or undetectable?? Tony give one picture showing the detectable status and undetectable status, keep an eye on lemmas (including proof)
    假设检验中，当概率足够小时拒绝 H_0, 多小是小？？ detection boundary 给出了指示

    2. Motivation for Higher Criticism (results from empirical process theory, see e.g. Shorack & Wellner(2009))

    3. HC 方法统计推断流程：计算 p 值；p 值标准化并排序；进行 HC 统计，当 HC 值过大时拒绝 H_0

    4. Testing procedures: linear statistic & linear test; scan statistic and scan test; final test

    5. 具体细节参见 Cai，T.T.，& W. （2017） Optimal screening and discovery of sparse with applications to multistage high-throughput studies. J. Roy. Statist. Soc. Ser. B.79

- Other References: Mentioned during class

[1] Detection: Ingster(1999) and Donoho & Jin (2004)

[2] Estimating the fraction p: Meinshausen & Rice (2006), Cai, Jin and Low(2006), Jin & Cai(2007)

[3] Detection boundary(Ingster, 1999; Donoho & Jin,2004

[4] Hellinger and total variation metrics (they are different)
Remarks: lemma of Inequalities relating Hellinger and total variation metrics

[5] Cai，Jeng & Jin（JRSSB，2011）consider the detection problem in the heteroscedastic case
Sparse case and dense case

【8.13】

- Theme
    1. Submatrix localization, see it in Cai, T.T., Liang, T. & Rakhlin, A.(2017) Computational and statistical boundaries for submatrix localization in a large noisy matrix. Ann. Statist. 45

    2. Robust sparse segment detection & identification: consider the noise distribution is not normal and unknown, see it in Cai, Jeng & Li(JRSSB,2012) considerer robust detection and identification when noise distribution is unknown and hard to be estimated.

3. Some interesting sparse signals detection problems:
   Detecting sparse Gaussian mixtures;
   Detecting of general (not necessarily Gaussian) mixtures
   Detecting submatrices
   Detecting Sparse segments in Gaussian noise
   Robust detection of sparse segments
   Detecting sparse segments with multiple samples
   Detecting principal components (PCA algorithm)

4. Testing a small number of hypothesis(Bonferroni method), testing a large number of hypotheses
5. FDR control(BH step-up procedure; proof of BH procedure);
   Improving BH procedure(Meinshausen & Rice(2006), Cai, Jin & Low(2007). Jin & Cai(2007). Cai & Jin(2010)
   Other FDR procedure:Adaptive procedure(Benjiamini and Hochberg,2000);Plug in procedure(Genovese and Wasserman,2004)
   FDR control based on p value, then give the proof of why we use p value as the evidence against H_0
6. Revisit Robbins(1951),see it in Robbins, H. (1951), Asymptotically subminimax solutions of compound statistical decision problems. Proc. 2nd Bereley Symp. Math. Statist. Probab. UC Berkeley
7. Data-Driven procedure: see it in
   http://www-stat.wharton.upenn.edu/~tcai/paper/html/FDR.html
8. Estimate p,f_0 and f with the desired properties. For estimating p and f0,see,Jin & Cai(2007) and Cai & Jin(2010);
   When all the assumptions are correct, the p values and z values follow their theoretical null distribution under the null
   The theoretical nulls could be quite different from the empirical nulls in applications. See Efron(2004) , Jin & Cai(2007), and Cai & Jin(2010)
   Uniformly consistent estimates of p and f0 are given in Jin & Cai(2007) and optimal rates are estimated in Cai & Jin(2010)
   Estimating f is a classical problem in nonparametric function estimation. See, e.g. Silverman(1986)

- Other references

[1] Efron. B.(2004) Large scale simultaneous hypothesis testing: The choice of a null hypothesis. J. Amer. Statist. Assoc 99,96

[2] Efron, B. (2008) . Microarrays, empirical Bayes and the two groups model . Statist. Sci. 23,1-22

[3] Benjamini, Y. & Yekutieli, D. (2001). The control of false discovery rate in multiple testing under dependency. Ann. Statist. 29

[4] Copas, J.(1974). On symmetric compound decision rules for dichotomies. Ann. Statist. 2,199-204

[5] Jin, J. & Cai, T.T. (2007). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. J. Amer. Statist. Assoc.102

[6] Cai, T.T. & Jin, J.(2010) Optimal rates of convergence for estimating the null and proportion of non-null effects in large-scale multiple testing. Ann. Statist. 38

[7] Cai, T.T.(2017) Global testing and large-scale multiple testing for high-dimensional covariance structures. Annu. Rev. Stat. Appl. 4

[8] Benjamini & Hochberg(1995,JRSSB): FDR 上界

【8.14】

- Theme
    1. Compressed sensing: Developing efficient algorithm to recover the unknown signal; Finding condition on A under which it is possible to recover beta accurately; constructing “good” sensing matrices A such that it is possible to recover sparse signals beta using an efficient algorithm
    2. Sparse signal recovery under MIP, see it in Cai, T.T., Wang, L. & Xu. G.(2010) Stable recovery of sparse signals and an oracle inequality. IEEE, Transactions on information Theory 56; Cai, T.T. & Zhang, A.(2013) Compressed sensing and affine rank minimization under restricted isometry, IEEE Trans. Signal Process 61; Cai, T.T. & Zhang, A. (2013). Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. IEEE Trans. Inf. Theory 60
    3. Understanding the relationships among the l_0(sparsity), l_1(objective function), l_2(loss) norms.
    4. Compressed sensing under RIP : different conditions on delta and theta have been used in the literature, see these in Candes and Tao(2005); Candes, Romberg and Tao(2006); Cades and Tao(2007); Candes(2008); Cai, Xu and Zhang(2009); Cai, Wang and Xu (2010); Cai and Zhang(2013a); Cai and Zhang(2013b)
    5. A variety of sufficient conditions on the RIC delta for the exact/stable recovery of k-sparse signals have been introduced in the literature. See these in Candes(2008); C., Wang and Xu(2012a); C., Wang and Xu(2010c)
    6. Coherence of random matrices: limiting laws, phase transition ,& application
    7. Problems about economics: see these in Andrews(Econometrica,1991) , Ligeralde and Brown (International Economic Review,1995)
- References

[1] Cai, T.T., Wang, L. & Xu, G. (2010) Stable recovery of sparse signals and an oracle inequality. IEEE Trans. Inf. Theory 56

[2] Cai, T.T., & Zhang, A. (2013) Compressed sensing and affine rank minimization under restricted isometry. IEEE Trans. Signal Process, 61

[3] Cai, T.T. & Zhang. A.(2014) Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. IEEE Trans. Inf. Theory 60

[4] Cai, T.T. & Jiang, T.(2011) Limiting laws of coherence of random matrices with application to testing covariance structure and construction of compressed sensing matrices. Ann. Statist. 39

[5] Baraniuk, R., Davenport, M., DeVore, R. & Wakin, M.(2008) A simple proof of the restricted isometry property for random matrices. Constr. Approx. 28

[6] Bickel, P. J., Ritov, Y. & Tsybakov, A. B. (2009) Simultaneous Analysis of Lasso and Dantzing Selector. Ann. Statist.

[7] Cai, T.T. & Zhang A.(2013) Sharp RIP bound for sparse signal and low-rank matrix discovery. Appl. Comput.

[8] Cai, T.T., Wang, L. & Xu, G.(2010). Shifting Ineuqality and recovery of sparse signals. IEEE Trans. Signal Process. 58

[9] Candes, E.J. & Tao, T. (2007) The Dantzing selector: statistical estimation when p is much larger than n(with discussion). Ann. Statist. 35

[10] DasGupta, S. & Gupta, A.(1999) An elementary proof of the Johnson-Lindenstrauss lemma

[11] Donoho, D.L. (2006). Compressed sensing. IEEE Trans. Inf. Theory 52

[12] Donoho, D.L., Elad, M. & Temlyakov, V.N. (2006) Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inf. Theory 52

[13] Johnson, W. B. & Lindenstrauss, J. (1984) Extension of Lipschitz mappings into a Hilbert space. Contemp. Math. 26

[14] Cades & Tao(2005,2007). Bickel, Ritov & Tsybakov(2008), Cai, Wang, & Xu(2010a,b), Cai and Zhang (2013a, b. 2014)

[15] Donoho & Huo(2001), Fuchs(2004,2005), Donoho, Elad & Temlyakov(2006) , Tropp(2004,2006), Cai, Wang & Xu(2010a, b) ,Cai and Zhang(2013a,b,2014)

[16] Cai, T.T., & Jiang. T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices, J. Multivariate Anal

Remarks:

Others you may interested about statistics:

Testing grouped hypothesis
Large-scale multiple testing with covariates
Integration large-scale data analysis & statistical inference
Statistical inference for high-dimensional linear regression
Inference for large matrices
Supervised learning
Unsupervised learning