Spring 2016

ECE598YW: Information-theoretic methods in high-dimensional statistics
Syllabus

Schedule:	Tuesday&Thursday 3:30-4:50pm, 3015 ECE Building
First lecture:	Tuesday, Jan 19 2016
Professor:	Yihong Wu yihongwu@illinois.edu, 129 CSL
Office hours:	Tuesday 11am-12pm or by appointment, 129 CSL
Website:	http://www.ifp.illinois.edu/~yihongwu/598.html

1 Content

The interplay between information theory and statistics is a constant theme in the development of both fields. This course will discuss how techniques rooted in information theory play a key role in understanding the fundamental limits of high-dimensional statistical problems in terms of minimax risk and sample complexity. In particular, we will rigorously justify the phenomena of dimensionality reduction by either "intrinsic low-dimensionality" (sparsity, smoothness, shape, etc) and - the less familiar – "extrinsic low-dimensionality" (functional estimation).

Complementing this objective of understanding the fundamental limits, another significant direction is to develop computationally efficient procedures that attain the statistical optimality, or to understand the lack thereof. Towards the end we will also discuss the recent trend of combining the statistical and algorithmic perspectives and the computational barriers in a series of statistical problems on large matrices and random graphs.

Tentative outline

1. **Introduction**: statistical experiment, decision-theoretic framework, minimax risk and Bayes risk, sample complexity, minimax theorem, least-favorable prior

2. **Review of information theory**: Entropy and mutual information, information-radius characterization of capacity, total variation, Hellinger distance, Kullback-Leibler (KL) divergence and operational characterizations, data processing principle, Pinsker inequality and joint range of f-divergences, Fisher information, Hammersley-Chapman-Robbins and Cramer-Rao lower bound.

3. **Metric entropy**: Covering and packing number, Gilbert-Varshamov bound and volumetric methods, Gaussian width, Sudakov minorization and Dudley integral, (*) Rate of convergence of least-squares estimators

4. **Unstructured estimation problems and oracle risk**: Gaussian location model (Warm-up), Anderson's lemma and exact minimax risk for subconvex loss, Minimax lower bounds via LeCam's method, Minimax lower bounds via Fano's inequality and Assouad's lemma, Examples: Denoising large matrices, Covariance matrix estimation and principle component analysis, Estimating distributions over large alphabets, Non-parametric model: Yang-Barron's characterization of minimax KL risk (example: density estimation)

5. **Structured estimation problem and excess risk**: Operator norms of random matrices, Sparse linear regression (identity and random design), Analysis of maximum likelihood (MLE) estimator, Convex relaxations of MLE: LASSO and Dantzig selector, Estimating structured covariance matrices, Inverse covariance matrix estimating and Gaussian graphical models.

6. **Functional testing and estimation**: LeCam's method revisited, simple-vs-composite and composite-vs-composite hypothesis testing, χ^2 divergence and the Ingster-Suslina technique, Estimating regression function/density at a point and general linear functionals, Testing high-dimensional covariance structures, Estimating non-smooth functions and polynomial approximation (ℓ_1 norm).

7. **Aggregation**: Aggregation, adaptivity, and oracle inequality, model selection, linear and convex aggregation

8. **High-dimensional sparse testing**: Testing sparse mixtures and detection boundary (Ingster-Donoho-Jin), Adaptive tests via empirical f-divergences.

9. **Advanced topics** (TBD): Statistical estimation on large graph,

   Statistical estimation on large alphabets, Algorithmic aspects of statistical inference.

10. **Final project presentation**

2 Administrivia

1. Course prerequisites: Maturity with probability theory at the level of ECE 534 or Math 561. ECE 563 will NOT be required. theory.

2. Final project: either presenting a paper or a standalone reseach project.

3. Grading: 30% attendence, 30% homeworks, 40% final project.

4. Materials: Lecture notes and additional reading materials will be posted online.

5. References:

   - I.A. Ibragimov and R.Z. Hasminskii. *Statistical Estimation: Asymptotic Theory*. Springer, 1981.
   - I.M. Johnstone. *Gaussian estimation: Sequence and wavelet models*, 2015. http://statweb.stanford.edu/~imj/GE09-08-15.pdf
   - A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.
   - A. Nemirovski. *Topics in non-parametric statistics*. In P. Bernard, editor, Ecole d'Eté de Probabilités de Saint-Flour 1998 volume XXVIII of Lecture Notes in Mathematics, New York: Springer, 2000. http://www2.isye.gatech.edu/~nemirovs/Lect_SaintFlour.pdf

- Pascal Massart. *Concentration Inequalities and Model Selection*. In J. Picard, editor, Ecole d'Eté de Probabilités de Saint-Flour 2003 volume XXXIII of Lecture Notes in Mathematics, New York: Springer, 2007.
- Sara van de Geer. Empirical Process Theory in M-Estimation, Cambridge, 2009. http://www.stat.math.ethz.ch/~geer/cowlas.pdf
- Y. Polyankiy and Y. Wu, Lecture notes on information theory, Feb 2015. http://www.ifp.illinois.edu/~yihongwu/teaching/itlectures.pdf