Spring 2024

S&DS: Topics in High-Dimensional Statistics and Information Theory
Syllabus

Schedule:	Tuesdays 4-550pm KT 207
Professor:	Yihong Wu yihong.wu@yale.edu
Office hours:	By appointment
Website:	http://www.stat.yale.edu/~yw562/teaching/SDS677

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) or – 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. **f-divergences**: total variation, Hellinger distance, Kullback-Leibler (KL) divergence, χ^2-divergence, and operational characterizations, data processing principle, Pinsker inequality and joint range of f-divergences, information bound (Fisher information, Hammersley-Chapman-Robbins, and Cramer-Rao lower bound)

3. **Variational representation**: Donsker-Varadhan, Gibbs variational formula, applications in ML (PAC Bayes generalization bound, variational autoencoder), applications in probability (concentration of sample covariance, small ball)

4. **Mutual information method (MIM)**: Entropy and mutual information, information-radius characterization of capacity, Anderson's lemma and exact minimax risk, lower bounds via MIM

5. **Minimax lower bounds**: Le Cam's two-point method, Assouad's lemma, Fano's inequality, connections to MIM. Examples: sparse estimation, community detection.

6. **Metric entropy**: Covering and packing number, Gilbert-Varshamov bound and volumetric methods, Gaussian width, Sudakov minorization and Dudley integral, Maurey's empirical method, metric entropy of smooth functions, connections to small-ball probability

7. **Entropic upper bounds for statistical estimation**: Yang-Barron's method for KL loss; universal compression and prediction. Composite hypothesis testing and Hellinger distance; Le Cam-Birgé's tournament for Hellinger loss. Yatracos' class and minimum distance estimator for TV loss. Example: Density estimation over Hölder classes.

8. **Functional estimation and testing**: Le Cam's method revisited, simple-vs-composite and composite-vs-composite hypothesis testing, χ^2 divergence and the second moment method (Ingster-Suslina). Examples: Estimating regression function/density at a point and general linear functionals, Testing high-dimensional covariance matrices, Estimating non-smooth functions and polynomial approximation (ℓ_1-norm). Testing sparse mixtures and detection boundary (Ingster-Donoho-Jin). Uniformity testing.

9. **Strong data processing inequality**: SDPI and SDPI constants, information percolation method. Examples: broadcasting on trees, spiked Wigner model, distributed correlation estimation.

10. **Advanced topics** (TBD):

    - Risk bound based on convex duality: dualizing two-point lower bound, compound decision and large alphabet problems.
    - Area theorem (I-MMSE identity) and statistical lower bound.
    - Aggregation: Aggregation, adaptivity, and oracle inequality, model selection, linear and convex aggregation

2 Administrivia

1. Course prerequisites: Maturity with probability theory. Prior exposure to information theory will **NOT** be required.

2. Final project: report on a research paper or a standalone research project.

3. Grading: 40% classroom participation, 30% homeworks, 30% final project.

4. Materials: We will draw materials from the following book draft

    - Y. Polyankiy and Y. Wu, *Information Theory: From Coding to Learning*, Cambridge University Press, forthcoming. http://www.stat.yale.edu/~yw562/teaching/itbook-export.pdf

    and some chapters from the following lecture notes

    - Y. Wu, *Information-theoretic Methods for High-dimensional Statistics*. http://www.stat.yale.edu/~yw562/teaching/it-stats.pdf

    Handout for other materials will be provided.

5. Other 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