Ziv Goldfeld
Office 322 Rhodes Hall
email: goldfeld@cornell.edu

Lectures: 2:55-4:10 PM T/TH
OH: 9-11 AM Wed.

Course Website: people.cornell.edu

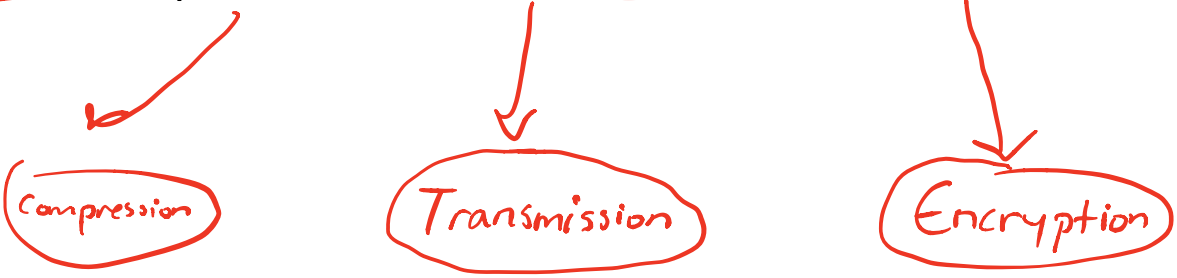
Grading:

HW - 1 every 2 weeks	20%
1 Midterm March 3	30%
1 Final Exam	50%

Information Theory: Mathematical framework for quantifying and rigorously reasoning about uncertainty and information (as a resolution of uncertainty)

+

Leverage this framework to study fundamental properties of "operations one can perform on information sources."



List of Topics

① Background: on Prob. Theory

② f-Divergences: Mathematical object capable of measuring the distance / proximity b/t prob. distributions

- $\mathcal{X}$ is a space
- $\mathcal{P}(\mathcal{X})$ is the set of all prob. measures on $\mathcal{X}$.
- f-divergence is a mapping

$$\delta : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}_{\geq 0}$$

$$\delta(P, Q) = 0 \iff P = Q$$

will cover KL-divergence, TV distance, χ^2 distance

③ Information Measures: Shannon Entropy, differentiable entropy, mutual information

④ Letter typical sequences

Let $\mathcal{X} = \{0, 1\}$ and define

$\mathcal{X}^n = \underbrace{\mathcal{X} \times \mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}}_{n\text{-fold Cartesian product}}$ = set of all binary sequences of length n

Note: $|\mathcal{X}^n| = |\mathcal{X}|^n \; 2^n$

Fix $\underline{P} \in \mathcal{P}(\{0,1\})$, namely $\underline{P} = \mathrm{Bern}(p)$, $p \in (0, 1/2)$

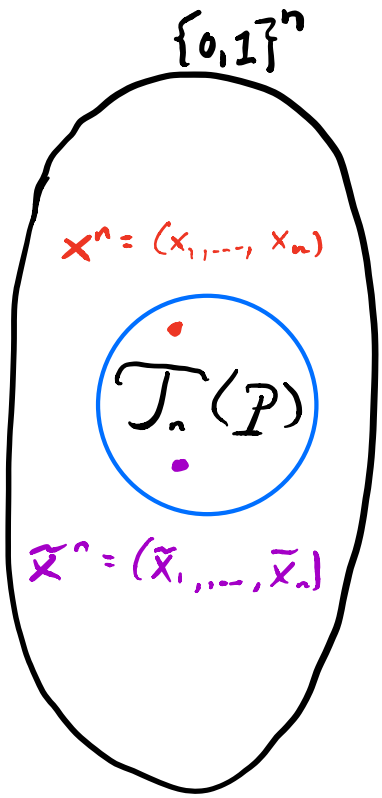
Let

$$X^n = (X_1, \ldots, X_n) \sim \underline{P}^{\otimes n}$$

One can define a subset of x^n

$$\mathcal{T}_n(\underline{P}) \subseteq x^n$$

such that



(i) $$\frac{|\mathcal{T}_n(\underline{P})|}{|x^n|} \xrightarrow[n\to\infty]{} 0$$

(ii) $$\mathbb{P}(X^n \in \mathcal{T}_n(\underline{P})) \xrightarrow[n\to\infty]{} 1$$

(iii) $$\forall\, x^n \in \mathcal{T}_n(\underline{P}): \quad \mathbb{P}(X^n = x^n) \approx \frac{1}{|\mathcal{T}_n(\underline{P})|}$$

(5) Reliable Communication over Noisy Channels

→ Operational setup, Shannon's channel coding Theorem.

proof $\begin{pmatrix} \text{direct:} & \text{typical sequences} \\ \text{converse:} & \text{prop. of info measures} \end{pmatrix}$

⑥ Distribution Simulation

- Fix $P \in \mathcal{P}(\mathcal{X})$
- Given iid $\text{Bern}(1/2)$ variables $Z_1, Z_2, \ldots$ design an algorithm $A(Z_1, Z_2, \ldots) \sim P$.

Approximate simulation

$$\delta(Q_A, P) \longrightarrow 0$$

⑦ Information-Theoretic Security

- Shannon's Cipher System (OTP)
- Wiretap Channel (Wyner 1975)
- Active wiretap channel

⑧ Information Theory and Machine Learning

- design ML algorithm for learning useful representations via info. bottleneck principle (Deep variation info bottleneck framework)
- use NNs + SGD to estimate MI (mutual information) (MINE: neural estimators)
- use adversarial learning to devise wiretap codes

I Background on Probability Theory

Probability Space: A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ such that

① Ω contains all possible outcomes

② $\mathcal{F}$ is a set of subsets of Ω, called the σ-field, satisfying

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c := \Omega / A \in \mathcal{F}$

(iii) $A_1, A_2, \ldots \in \mathcal{F} \rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

③

… Next time