

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."

Compression

Transmission

Encryption

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

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

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

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

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

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

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

Let

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

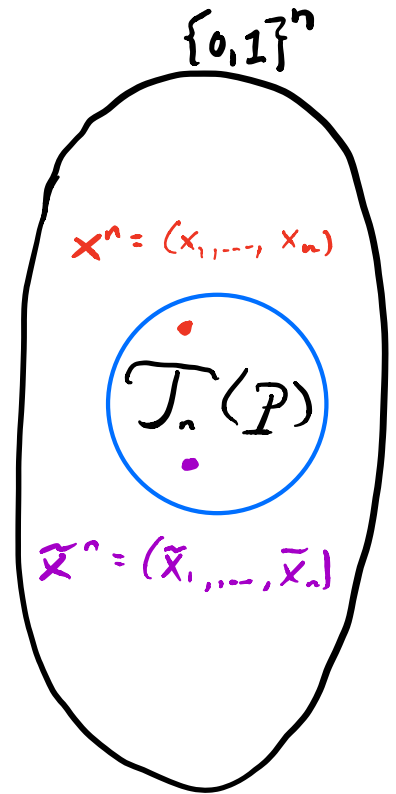
One can define a subset of X^n

$$\mathcal{T}_n(\underline{p}) \subseteq X^n$$

such that

$$(i) \quad \frac{|\mathcal{T}_n(\underline{p})|}{|X^n|} \xrightarrow{n \rightarrow \infty} 0$$

$$(ii) \quad \mathbb{P}(X^n \in \mathcal{T}_n(\underline{p})) \xrightarrow{n \rightarrow \infty} 1$$



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

⑤ Reliable Communication over Noisy Channels

→ Operational setup, Shannon's channel coding Theorem,

proof $\left(\begin{array}{l} \text{direct:} \quad \text{typical sequences} \\ \text{converse:} \quad \text{prop. of info measures} \end{array} \right)$

⑥ Distribution Simulation

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

Approximate simulation

$$\delta(Q, P) \rightarrow D$$

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 $\overset{\text{mutual information}}{\downarrow} \text{MI}$ (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) $A \in \mathcal{F} \Rightarrow A^c := \Omega \setminus A \in \mathcal{F}$

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

③

... Next time