ECE253/CSE208 Introduction to Information Theory

Lecture 16: Summary

Dr. Yu Zhang

ECE Department
University of California, Santa Cruz

- Chap 1–10 of *Elements of Information Theory (2nd Edition) by Thomas Cover & Joy Thomas*.

What We Have Covered

Key Concepts

- Entropy and differential entropy; KL divergence; mutual information.
- DPI; sufficient statistics; convex functions; Jensen's ineq; Fano's ineq.
- AEP; typical set; entropy rate; Markov chain; stationary stochastic process.
- Source coding theorem; Kraft-McMillan inequality; competitive optimality.
- Lossless source coding: Huffman, Shannon, SF, SFE, and arithmetic coding.
- Horse race; log wealth relative; double rate; financial value of side information.
- Channel coding theorem and its proof; channel capacity of "simple" channels.
- Channel capacity of AWGN/parallel AWGN/feedback; water-filling power allocation.
- Rate-distortion theory; Blahut–Arimoto algorithm.

For Discrete Distributions

PMF $p(x)$	Entropy
Definition	$H(X) = -\sum_x p(x) \log p(x)$
Bounds	$[0, \log \lvert\mathcal{X}\rvert]$
$H_{\max}$ distribution	Uniform
Translation	$H(X+c) = H(X)$
Scaling	$H(cX) = H(X)$
Joint entropy	$H(X,Y) = -\mathrm{E}_{(X,Y)} \log p(X,Y)$
Conditional entropy	$H(Y\lvert X) = -\mathrm{E}_{(X,Y)} \log p(Y\lvert X)$
Relative entropy	$D(p\lvert\lvert q) = \sum p \log \frac{p}{q}$
Mutual information	$I(X;Y) = D\big(p(X,Y)\lvert\lvert p(X)p(Y)\big)$
Chain rule	$H(X^n) = \sum_{i=1}^n H(X_i\lvert X^{i-1})$
AEP (i.i.d. X^n)	$-\frac{1}{n} \log p(X^n) \xrightarrow{\text{i.p.}} H(X)$
Typical set	$A_\epsilon^{(n)} = \left\{x^n : \lvert -\frac{1}{n} \log p(x^n) - H(X)\rvert \le \epsilon\right\}$

For Continuous Distributions

PDF $p(x)$	Differential Entropy
Definition	$h(X) = -\int_{\mathcal{S}} f(x) \log f(x)\, dx$
Bounds	$(-\infty, \frac{1}{2}\log(2\pi e\sigma^2)]$
$H_{\max}$ distribution	Gaussian
Translation	$h(X+c) = h(X)$
Scaling	$h(cX) = h(X) + \log\|c\|$
Joint entropy	$h(X,Y) = -\mathrm{E}_{(X,Y)} \log f(X,Y)$
Conditional entropy	$h(Y\|X) = -\mathrm{E}_{(X,Y)} \log f(Y\|X)$
Relative entropy	$D(f\|\|g) = \int f \log(\frac{f}{g})$
Mutual information	$I(X;Y) = D\big(f(x,y)\|\|f(x)f(y)\big)$
Chain rule	$h(X^n) = \sum_{i=1}^{n} h(X_i\|X^{i-1})$
AEP (i.i.d. X^n)	$-\frac{1}{n}\log f(X^n) \xrightarrow{\text{i.p.}} h(X)$
Typical set	$A_\epsilon^{(n)} = \{x^n : \|-\frac{1}{n}\log f(x^n) - h(X)\| \le \epsilon\}$

Typical Sequences and Set

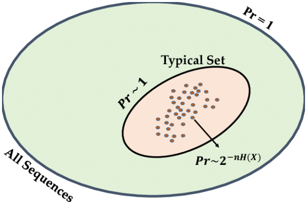

Figure: For n sufficiently large, all typical sequences have about the same probability $2^{nH(X)}$ (*asymptotic equipartition*). Everything outside the typical set has a negligible probability.

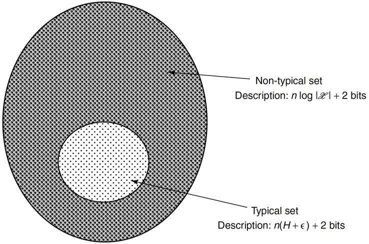

Figure: Encoding for the typical set: On average $H(X)$ bits are needed to encode X^n per symbol.

Lossless Source Coding

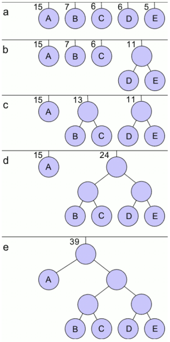

Fig 1. Huffman tree (bottom-up)

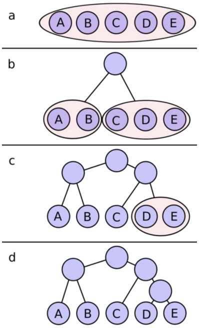

Fig 2. SF tree (top-down)

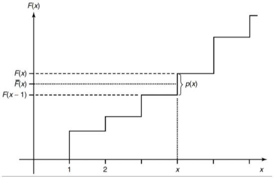

Fig 3. SFE coding

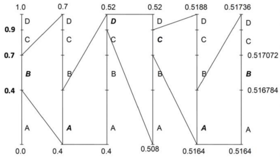

Fig 4. Arithmetic coding

Channel Coding Theorem

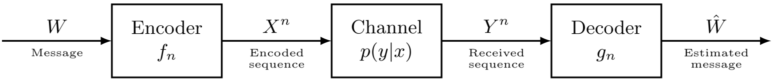

Figure: Channel capacity C is the sharp threshold between reliable and unreliable communication.

- *Information* channel capacity: $C = \max_{p(x)} I(X;Y)$
- *Operational* channel capacity: number of bits transmitted reliably (with an arbitrarily small probability of error) per channel usage.
- Information channel capacity $=$ operational channel capacity.
- **Achievability**: All rates below capacity $R < C$ are achievable.
- **Converse**: $(2^{nR}, n)$ code with probability of error $\lambda^{(n)} \xrightarrow{n\to\infty} 0$ must have $R \leq C$.

Shannon Limit of Channel Capacity

$$\boxed{C_{\mathsf{AWGN}} = W \log_2\left(1 + \frac{P}{N_0 W}\right)}$$

$$\max_{P_i \geq 0} \quad \sum_{i=1}^{N} \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) \tag{1a}$$

$$\text{s.t.} \quad \sum^{N} P_i = P \tag{1b}$$

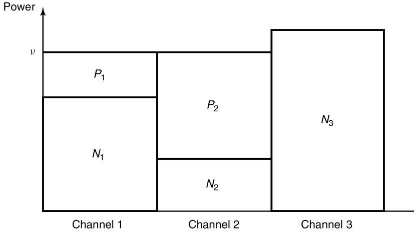

Figure: Water-filling optimal power allocation: **allocate more power in less noisy channels**. The optimal solution is obtained as: $P_i^* = (\nu - N_i)^+, \ i = 1, 2, \ldots k$, where the water level ν satisfies $\sum_i (\nu - N_i)^+ = P$.

Rate Distortion Coding Theorem

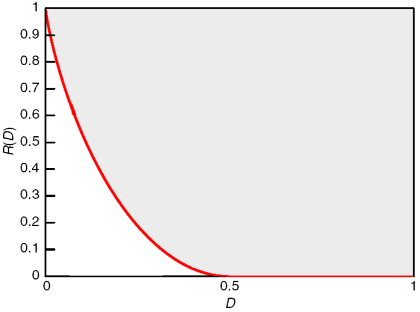

Figure: The rate-distortion function of a Bernoulli random variable with Hamming distortion.

Theorem (Rate Distortion Coding Theorem)

The rate-distortion function for an i.i.d. source X with distribution $p(x)$ and bounded distortion function $d(x, \hat{x})$ is equal to the associated information rate distortion function. That is, $R(D) = R^{(I)}(D)$.

What We Have Not Covered

Network Information Theory[1]

- "A system with many senders and receivers contains many new elements in the communication problem: **interference, cooperation, and feedback**. These are the issues that are the domain of network information theory. Given many senders and receivers and a channel transition matrix that describes the effects of the interference and the noise in the network, decide whether or not the sources can be transmitted over the channel....This general problem has not yet been solved...."

- The capacity of a general wireless network is not known; see `https://en.wikipedia.org/wiki/List_of_unsolved_problems_in_information_theory`

[1] *Network Information Theory*, A. Gamal and Y. Kim, Cambridge University Press (2011).

Network Information Theory (cont'd)

- Broadcast Channels (one-to-many communication)

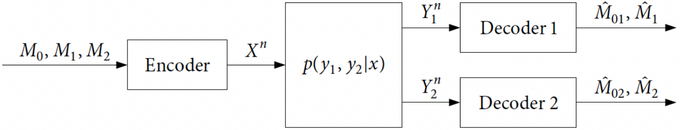

Figure 5.1. Two-receiver broadcast communication system.

Network Information Theory (cont'd)

- Multiple Access Channels (many-to-one communication)

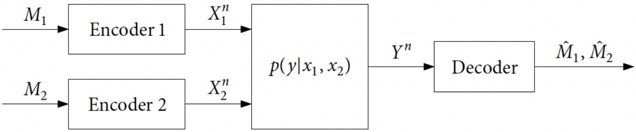

Figure 4.1. Multiple access communication system with independent messages.

Network Information Theory (cont'd)

- Multiple-Input and Multiple-Output (MIMO) channels: using multiple transmission and receiving antennas to exploit multipath propagation

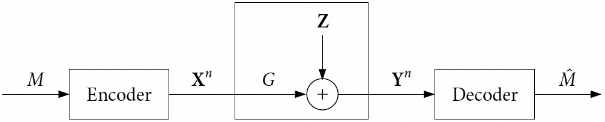

Figure 9.1. MIMO point-to-point communication system.

Network Information Theory (cont'd)

- Graphical Networks: multi-hop networks, where some nodes can act as both senders and receivers and hence communication can be performed over multiple rounds.

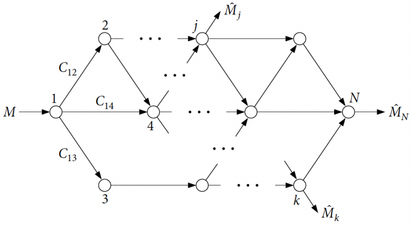

Figure 15.1. Graphical multicast network.

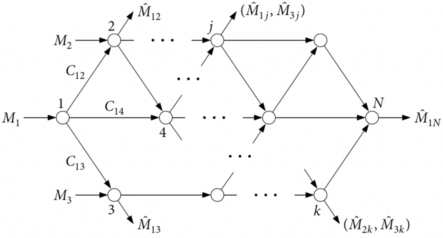

Figure 15.9. Graphical multimessage network.

Network Information Theory (cont'd)

- Relay Channels, e.g., communication between two base stations through both a terrestrial link and a satellite, or between two nodes in a mesh network with an intermediate node acting as a relay.
- The capacity of the relay channel is not known in general.

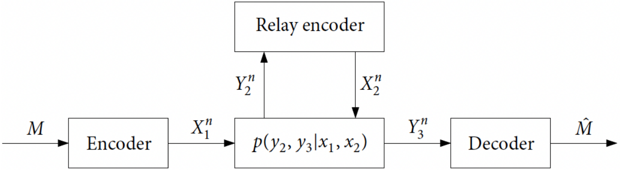

Figure 16.1. Point-to-point communication system with a relay.

Thank You!

Email: <zhangy@ucsc.edu>

Homepage: https://people.ucsc.edu/~yzhan419/