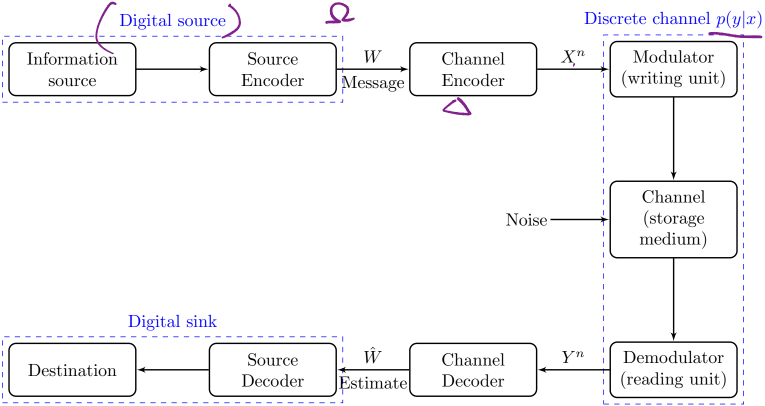
ECE 231A Discussion 4

TA: Hengjie Yang

Email: hengjie.yang@ucla.edu

04/24/2020

Discrete channel



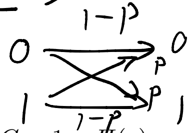
Discrete channel: a system consisting of input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a probability transition matrix $p(y|x)$.

Memoryless: $p(y_n|x_n, x^{n-1}, y^{n-1}) = p(y_n|x_n), \forall n \in \mathbb{N}, \forall (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}$

Capacity of discrete memoryless channels (DMCs)

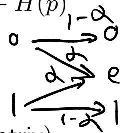
(Information) channel capacity: the information channel capacity for a discrete memoryless channel (DMC) is defined as

$$
\begin{aligned}
C &\triangleq \max_{p(x), x\in\mathcal{X}} I(X;Y) \\
&= \max_{p(x), x\in\mathcal{X}} \sum_{x\in\mathcal{X}} p(x) \left(\sum_{y\in\mathcal{Y}} p(y|x) \log \frac{p(y|x)}{\sum_{x'\in\mathcal{X}} p(x')p(y|x')} \right) \\
&= \max_{p(x), x\in\mathcal{X}} \sum_{x\in\mathcal{X}} p(x) D(P(Y|X=x)\|P(Y))
\end{aligned}
$$



Examples:

1. Binary symmetric channel (BSC) with crossover prob. p: $C = 1 - H(p)$ with $p^*(x) = 1/2, x \in \{0,1\}$.



2. Binary erasure channel (BEC) with erasure prob. α: $C = 1 - \alpha$ with $p^*(x) = 1/2, x \in \{0,1\}$.
3. Weakly symmetric channel: $C = \log|\mathcal{Y}| - H(\text{row of transition matrix})$ with $p^*(x) = 1/|\mathcal{X}|, x \in \mathcal{X}$.
4. Cyclic symmetric channel: C is achieved with $p^*(x) = 1/|\mathcal{X}|, x \in \mathcal{X}$.

Properties of channel capacity and KKT conditions

Theorem:

$$0 \leq C \leq \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}.$$

1. $I(X;Y)$ is concave of $p(x)$ over a closed convex set for a fixed $p(y|x)$. Hence, the maximum is finite and unique.
2. In general, there is no closed-form solution to capacity.
3. Capacity can be found efficiently using Blahut-Arimoto algorithm.

KKT conditions: $p(x), x \in \mathcal{X}$ is the capacity-achieving distribution for $I(X;Y)$ if for some constant C,

$$\max_{p(x)} I(X;Y)$$

$$\text{s.t. } \sum_{x \in \mathcal{X}} p(x) = 1$$

$$D\big(P(Y|X=x)\|P(Y)\big) = C, \quad \text{if } p(x) > 0$$
$$D\big(P(Y|X=x)\|P(Y)\big) \leq C, \quad \text{if } p(x) = 0$$

Furthermore, C is the capacity of the channel.

$$L^{*}(p(x), \lambda) = I(X;Y) - \lambda\left(\sum p(x) - 1\right)$$

Proof: Take the partial derivative $\frac{\partial I(X;Y)}{\partial p(x)} = D\big(P(Y|X=x)\|P(Y)\big) - \log e$ and then apply the KKT conditions.

$$\frac{\partial L}{\partial p(x)} = \boxed{\frac{\partial I(X;Y)}{\partial p(x)}} - \lambda \overset{\Delta}{=} 0$$

$$\leq$$

Operational definition of capacity

$$\xrightarrow[\text{Message}]{W} \boxed{\text{Encoder}} \xrightarrow{X^n} \boxed{\begin{array}{c}\text{Channel}\\ p(y|x)\end{array}} \xrightarrow{Y^n} \boxed{\text{Decoder}} \xrightarrow[\text{Estimated message}]{\hat{W}}$$

Setup:

M: # of msgs

n: # of channel uses

1. $\Omega = \{1, 2, \ldots, M\}$: an index set
2. $W \in \Omega$: a message drawn from Ω
3. $(\mathcal{X}, p(y|x), \mathcal{Y})$: the discrete channel, with $\sum_{y \in \mathcal{Y}} p(y|x) = 1, x \in \mathcal{X}$.
4. $(\mathcal{X}^n, p(y|x), \mathcal{Y}^n)$: the n-th extension of DMC, $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$
5. An (M, n) code for channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of "codebook"
   (i) An index set $\Omega = \{1, 2, \ldots, M\}$;
   (ii) An encoding function $X^n : \Omega \to \mathcal{X}^n$, yielding codewords $x^n(1), \ldots, x^n(M)$
   (iii) A decoding function $g : \mathcal{Y}^n \to \Omega$, a deterministic rule which assigns a guess to each y^n
6. $\lambda_i \triangleq \Pr(g(Y^n) \neq i | X^n = x^n(i))$: condi. prob. of error by sending index i
7. $\lambda^{(n)} \triangleq \max_{i \in \Omega} \lambda_i$: the maximal prob. of error for an (M, n) code
8. $P_e^{(n)} \triangleq \frac{1}{M} \sum_{i=1}^M \lambda_i$: the average prob. of error for an (M, n) code
9. $R \triangleq \frac{\log M}{n}$: the rate of an (M, n) code.
10. $C = \sup\{R : \exists(\lceil 2^{nR} \rceil, n) \text{ codes with } \lambda^{(n)} \to 0\}$ (Operational definition).

Jointly typical sets and joint AEP

Jointly typical set $A_\epsilon^{(n)}$: the set of sequences $\{(x^n, y^n)\}$ w.r.t. $p(x,y)$:

$$A_\epsilon^{(n)} = \Bigg\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| -\frac{\log p(x^n)}{n} - H(X) \right| < \epsilon, \\ \left| -\frac{\log p(y^n)}{n} - H(Y) \right| < \epsilon, \\ \left| -\frac{\log p(x^n, y^n)}{n} - H(X,Y) \right| < \epsilon \Bigg\}$$

where $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$.

Joint AEP: Let (X^n, Y^n) be sequences of length-n drawn i.i.d. according to $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$, then

(i) for a given $\epsilon > 0$, $\lim_{n\to\infty} \Pr\left\{ (X^n, Y^n) \in A_\epsilon^{(n)} \right\} = 1$;

(ii) $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$, and $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X,Y)-\epsilon)}$ for sufficiently large n;

(iii) If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, then $\Pr\left\{ (\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)} \right\} \leq 2^{-n(I(X;Y)-3\epsilon)}$; and $\Pr\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)} \right) \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}$ for sufficiently large n.

Channel coding theorem cont.'d

Channel coding theorem: For a DMC, all rates below capacity C are achievable. Specifically, for every $R < C$, there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes with maximum prob. of error $\lambda^{(n)} \to 0$. Conversely, any sequence of $(\lceil 2^{nR} \rceil, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \le C$.

$C = \max I(X;Y)$ $\qquad p(x^n) = \prod_{i=1}^{n} p(x_i)$

Outline of the proof:

1. Fix some $p^*(x)$, generate $(\lceil 2^{nR} \rceil, n)$ code at random $\sim p^*(x)$.
2. A message W is chosen according to uniform distribution. This leads to
   ave. prob. of error $P_e^{(n)} = \Pr\{g(Y^n) \neq W\}$. error prob.
3. The decoder employs jointly typical decoding: $\hat{W}$ is declared if
   - (i) $(X^n(\hat{W}), Y^n)$ is jointly typical
   - (ii) No other $W' \neq \hat{W}$ exists s.t. $(X^n(W'), Y^n) \in A_\epsilon^{(n)}$.

   If no such $\hat{W}$ exists, or more than one such, an error is declared.

   $p(X^n(1))$
   $p(X^n(2))$
   $p(X^n(2^{nR}))$
4. Let $\mathcal{E} \triangleq \{\hat{W} \neq W\}$ and $E_i \triangleq \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\}$. Then $\quad \Pr(\mathcal{E}) = \Pr(\mathcal{E} | W=1)$

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E}|W=1) = \Pr(E_1^c \cup E_2 \cup \cdots \cup E_{2^{nR}} | W=1)$$

$\Pr\left((X^n(1), Y^n) \notin A_\epsilon^{(n)} \right) \to 1$ $\qquad (X^n(w'), Y^n) \in A_\epsilon^{(n)}$

$$\le \Pr(E_1^c | W=1) + \sum_{i=2}^{2^{nR}} \Pr(E_i | W=1)$$

$w' \neq 1$, $\le 2^{-n(I(X;Y) - 3\epsilon)}$

$R < I(X;Y)$, $p(x)$

$$\le \epsilon + 2^{-n(I(X;Y) - 3\epsilon - R)} \qquad (2^{nR} - 1) \le 2$$

$\le \epsilon$ $\qquad R \to R - \frac{1}{n}$

$$\le 2\epsilon \quad (n \text{ large enough and } R < I(X;Y) = C)$$

Exercise: Channels with memory have higher capacity

Consider a BSC with $Y_i = X_i \oplus Z_i$, where $\oplus$ is a mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $Pr\{Z_i = 1\} = p$, but that $Z_1, \ldots, Z_n$ are not necessarily independent. Assume Z^n is independent of X^n. Let $C = 1 - H(p)$. Show that

$$\max_{p(x_1,\ldots,x_n)} I(X_1, \ldots, X_n; Y_1, \ldots, Y_n) \geq nC.$$

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n)$$

$$H(X^n \oplus Z^n | X^n) = H(Z^n|X^n) = H(Z^n) = \sum_{i=1}^{n} H(Z_i|Z^{i-1}) \leq \sum_{i=1}^{n} \underbrace{H(Z_i)}_{H(p)}$$

$$\begin{aligned}
I(X^n;Y^n) &= H(X^n) - H(X^n|Y^n) \\
&= H(X^n) - H(Y^n \oplus Z^n | Y^n) \\
&= H(X^n) - \underbrace{H(Z^n|Y^n)}_{\leq H(Z^n)} \\
&\geq H(X^n) - H(Z^n) \\
&= H(X^n) - \sum_{i=1}^{n} H(Z_i|Z^{i-1}) \\
&\geq H(X^n) - \sum_{i=1}^{n} H(Z_i) \\
&\geq H(X^n) - nH(p)
\end{aligned}$$

X^n is iid $\sim \mathrm{Bern}(\frac{1}{2})$

$$\max I(X^n;Y^n) \geq H(X^n)\Big|_{\substack{p(x_i)=\frac{1}{2} \\ X^n \text{ iid}}} - nH(p) = n - nH(p) = nC.$$