ECE 587 Midterm II Review

Miao Liu

Department of Electrical and Computer Engineering
Duke University, Durham NC 27708

November 5, 2012

- **Channel capacity.** The logarithm of the number of distinguishable inputs is given by

$$C = \max_{p(x)} I(X; Y) \tag{1}$$

- **Channel capacity.** The logarithm of the number of distinguishable inputs is given by

$$C = \max_{p(x)} I(X; Y) \tag{1}$$

- **Examples**
  - Binary symmetric channel: $C = 1 - H(p)$
  - Binary erasure channel: $C = 1 - \alpha$
  - Symmetric channel:
    $C = log|\mathcal{Y}| - H(\text{row of transition matrix})$

- **Channel capacity.** The logarithm of the number of distinguishable inputs is given by

$$C = \max_{p(x)} I(X;Y) \tag{1}$$

- **Examples**
  - Binary symmetric channel: $C = 1 - H(p)$
  - Binary erasure channel: $C = 1 - \alpha$
  - Symmetric channel:
    $C = log|\mathcal{Y}| - H(\text{row of transition matrix})$
- **Properties of C**
  1. $0 \leq C \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$
  2. $I(X;Y)$ is a continuous concave function of $p(x)$

- **Joint typicality.** The set $A_\epsilon^{(n)}$ of joint typical sequences $\{(x^n, y^n)\}$ w.r.t. the distribution $p(x, y)$ is given by

$$
\begin{aligned}
A_\epsilon^{(n)} = \Big\{ & (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \\
& \big| -\frac{1}{n} p(x^n) - H(X) \big| \leq \epsilon, \\
& \big| -\frac{1}{n} p(y^n) - H(Y) \big| \leq \epsilon, \\
& \big| -\frac{1}{n} p(x^n, y^n) - H(X, Y) \big| \leq \epsilon \Big\},
\end{aligned}
\tag{2}
$$

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

- **Joint typicality.** The set $A_\epsilon^{(n)}$ of joint typical sequences $\{(x^n, y^n)\}$ w.r.t. the distribution $p(x, y)$ is given by

$$A_\epsilon^{(n)} = \Big\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ \big| -\frac{1}{n} p(x^n) - H(X) \big| \leq \epsilon, \\ \big| -\frac{1}{n} p(y^n) - H(Y) \big| \leq \epsilon, \\ \big| -\frac{1}{n} p(x^n, y^n) - H(X, Y) \big| \leq \epsilon \Big\}, \tag{2}$$

  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:
  1. $Pr((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$ as $n \to \infty$.
  2. $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$.
  3. If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, then
     $Pr\Big((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)} \Big) \leq 2^{-n(I(X;Y)-3\epsilon)}$.

- **Channel coding theorem.** All rates below capacity C are achievable, and all rates above capacity are not; that is, for all rates $R < C$, there exists a sequence of $(2nR, n)$ codes with probability of error$\lambda^{(n)} \to 0$. Conversely, for rates $R > C$, $\lambda^{(n)}$ is bounded away from 0.

- **Channel coding theorem.** All rates below capacity C are achievable, and all rates above capacity are not; that is, for all rates $R < C$, there exists a sequence of $(2nR, n)$ codes with probability of error$\lambda^{(n)} \to 0$. Conversely, for rates $R > C$, $\lambda^{(n)}$ is bounded away from 0.
- **Feedback capacity.** Feedback does not increase capacity for discrete memoryless channels (i.e., $C_{FB} = C$).

- **Channel coding theorem.** All rates below capacity C are achievable, and all rates above capacity are not; that is, for all rates $R < C$, there exists a sequence of $(2nR, n)$ codes with probability of error$\lambda^{(n)} \to 0$. Conversely, for rates $R > C$, $\lambda^{(n)}$ is bounded away from 0.
- **Feedback capacity.** Feedback does not increase capacity for discrete memoryless channels (i.e., $C_{FB} = C$).
- **Source-Channel theorem.** A stochastic process with entropy rate H cannot be sent reliably over a discrete memoryless channel if $H > C$. Conversely, if the process satisfies the AEP, the source can be transmitted reliably if $H < C$.

HW6-3

Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of the channel.

HW6-3

Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of the channel.

Solution: the questions is equivalent to find the distribution $p(x_1, x_2)$ that maximizes $I(X_1, X_2; Y_1, Y_2)$.

HW6-3

Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of the channel.

Solution: the questions is equivalent to find the distribution $p(x_1, x_2)$ that maximizes $I(X_1, X_2; Y_1, Y_2)$.
Since $p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)$, therefore, we have $Y_1 \leftarrow X_1 \leftrightarrow X_2 \rightarrow Y_2$ and

HW6-3

Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of the channel.

Solution: the questions is equivalent to find the distribution $p(x_1, x_2)$ that maximizes $I(X_1, X_2; Y_1, Y_2)$.
Since $p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)$, therefore, we have $Y_1 \leftarrow X_1 \leftrightarrow X_2 \rightarrow Y_2$ and

$$
\begin{aligned}
I(X_1, X_2; Y_1, Y_2) =& H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) \\
=& H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) && (3) \\
=& H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (4) \\
\leq& H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (5) \\
=& I(X_1, Y_1) + I(X_2, Y_2) && (6)
\end{aligned}
$$

HW6-3 continues

Therefore

$$C = \max_{p(x_1, x_2)} I(X_1, X_2; Y_1, Y_2)$$

HW6-3 continues

Therefore

$$
\begin{aligned}
C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) \\
&\leq \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2)
\end{aligned}
$$

HW6-3 continues

Therefore

$$
\begin{aligned}
C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) \\
&\leq \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) \\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2)
\end{aligned}
$$

HW6-3 continues

Therefore

$$
\begin{aligned}
C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) \\
&\leq \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) \\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) \\
&= C_1 + C_2
\end{aligned}
$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^(x_2)$ and $p*(x_1)$ and $p*(x_2)$ are the distributions that maximize C_1 and C_2 respectively.

HW7-2

Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $p(Z_i = 1) = p = 1 - P(Z_i = 0)$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Let $C = 1 - H(p)$. Show that $\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC$. Comment on the implications.

HW7-2

Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $p(Z_i = 1) = p = 1 - P(Z_i = 0)$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Let $C = 1 - H(p)$. Show that $\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC$. Comment on the implications.

Solution: When $X_1, X_2, \cdots, X_n$ are chosen i.i.d from $Bern(\frac{1}{2})$, Let $X^{(n)} = \{X_1, X_2, \cdots, X_n\}$ and $Y^{(n)} = \{Y_1, Y_2, \cdots, Y_n\}$

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

HW7-2

Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $p(Z_i = 1) = p = 1 - P(Z_i = 0)$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Let $C = 1 - H(p)$. Show that $\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC$. Comment on the implications.

Solution: When $X_1, X_2, \cdots, X_n$ are chosen i.i.d from $Bern(\frac{1}{2})$, Let $X^{(n)} = \{X_1, X_2, \cdots, X_n\}$ and $Y^{(n)} = \{Y_1, Y_2, \cdots, Y_n\}$

$$\begin{aligned} I(X^{(n)}; Y^{(n)}) &= H(X^{(n)}) - H(X^{(n)} | Y^{(n)}) \\ &= H(X^{(n)}) - H(Z^{(n)} | Y^{(n)}) \end{aligned}$$

HW7-2

Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $p(Z_i = 1) = p = 1 - P(Z_i = 0)$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Let $C = 1 - H(p)$. Show that $\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC$. Comment on the implications.

Solution: When $X_1, X_2, \cdots, X_n$ are chosen i.i.d from $Bern(\frac{1}{2})$, Let $X^{(n)} = \{X_1, X_2, \cdots, X_n\}$ and $Y^{(n)} = \{Y_1, Y_2, \cdots, Y_n\}$

$$\begin{aligned}
I(X^{(n)}; Y^{(n)}) &= H(X^{(n)}) - H(X^{(n)} | Y^{(n)}) \\
&= H(X^{(n)}) - H(Z^{(n)} | Y^{(n)}) \\
&\geq H(X^{(n)}) - H(Z^{(n)})
\end{aligned}$$

HW7-2

Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $p(Z_i = 1) = p = 1 - P(Z_i = 0)$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Let $C = 1 - H(p)$. Show that $\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC$. Comment on the implications.

Solution: When $X_1, X_2, \cdots, X_n$ are chosen i.i.d from $Bern(\frac{1}{2})$, Let $X^{(n)} = \{X_1, X_2, \cdots, X_n\}$ and $Y^{(n)} = \{Y_1, Y_2, \cdots, Y_n\}$

$$\begin{aligned}
I(X^{(n)}; Y^{(n)}) &= H(X^{(n)}) - H(X^{(n)}|Y^{(n)}) \\
&= H(X^{(n)}) - H(Z^{(n)}|Y^{(n)}) \\
&\geq H(X^{(n)}) - H(Z^{(n)}) \\
&\geq H(X^{(n)}) - \sum_{i=1}^{n} H(Z_i) \\
&= nH(1/2) - nH(p)
\end{aligned}$$

HW7-2 continues

Therefore, with memory over n uses of the channel, the channel capacity is

$$
\begin{aligned}
nC^{(n)} &= \max_{p(X^n)} I(X^{(n)}; Y^{(n)}) \\
&\geq I(X^{(n)}; Y^{(n)}) \\
&\geq n(1 - H(p)) \\
&= nC
\end{aligned}
$$

HW7-2 continues

Therefore, with memory over n uses of the channel, the channel capacity is

$$\begin{aligned} nC^{(n)} &= \max_{p(X^n)} I(X^{(n)}; Y^{(n)}) \\ &\geq I(X^{(n)}; Y^{(n)}) \\ &\geq n(1 - H(p)) \\ &= nC \end{aligned}$$

- Conclusion: channels with memory have higher capacity.

HW7-2 continues

Therefore, with memory over n uses of the channel, the channel capacity is

$$\begin{aligned} nC^{(n)} &= \max_{p(X^n)} I(X^{(n)}; Y^{(n)}) \\ &\geq I(X^{(n)}; Y^{(n)}) \\ &\geq n(1 - H(p)) \\ &= nC \end{aligned}$$

- Conclusion: channels with memory have higher capacity.
- Intuition
  - correlation between the noise decreases the effective noise.
  - the information from the past samples of noise helps reducing the present noise.

Summary

Good Luck!