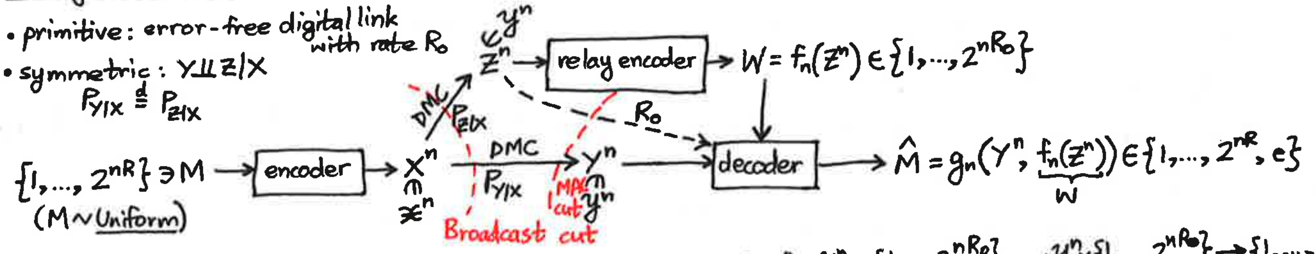
Anuran Makur
5 Dec 2016

INFORMATION INEQUALITY:

① Relay Channel:

- primitive: error-free digital link with rate R_0
- symmetric: $Y \perp\!\!\!\perp Z | X$
  $P_{Y|X} \overset{d}{=} P_{Z|X}$



- $\underline{(2^{nR}, n)\text{-code}}$: (message set $\{1,\dots,2^{nR}\}$, $\underset{\uparrow \text{ encoder}}{x^n}: \{1,\dots,2^{nR}\} \to \mathcal{X}^n$, $\underset{\uparrow \text{ relay encoder}}{f_n}: \mathcal{Y}^n \to \{1,\dots,2^{nR_0}\}$, $\underset{\uparrow \text{ decoder}}{g_n}: \mathcal{Y}^n \times \{1,\dots,2^{nR_0}\} \to \{1,\dots,2^{nR}, e\}$)
- Probability of error: $P_e^{(n)} = \mathbb{P}(\hat{M} \neq M)$ ← average
- Rate R is achievable if $\exists$ sequence of $(2^{nR}, n)$-codes with $P_e^{(n)} \to 0$ as $n \to \infty$.
- Capacity: $C(R_0) \triangleq \sup\{R \geq 0 : R \text{ achievable}\}$.

② Cutset Bound: (Cover-ElGamal '79)

choice of codes determine P_{X^n} from P_M via $x^n(\cdot)$

For a primitive relay channel, if rate R is achievable, then $\exists P_X$ such that:

construct from P_{X_i}'s by averaging

$$R \leq I(X; Y, Z) \,, \quad \text{[Broadcast bound]}$$
$$R \leq I(X; Y) + R_0 \,. \quad \text{[MAC bound]}$$

Remarks:

Tightness

1. $C(0) = \max_{P_X} I(X;Y)$ and $C(\infty) = \max_{P_X} I(X;Y,Z)$. [Cover: Find min R_0 s.t. $C(R_0) = \max_{P_X} I(X;Y,Z)$.]

   ↑ MAC equality ↑ Broadcast equality (true for $R_0 \geq \log(|\mathcal{Y}|)$)

2. Reverse "Physically" Degraded Primitive Channel: (not symmetric)

   $P_{Y,Z|X} = P_{Y|X} P_{Z|Y}$ $[Z \perp\!\!\!\perp X | Y]$ $X \to Y \to Z$

   $I(X; Y, Z) = H(X) - H(X|Y,Z) = H(X) - H(X|Y) = I(X;Y)$

   $\Rightarrow R \leq I(X;Y)$ [Broadcast bound]

   So, $R = \max_{P_X} I(X;Y)$ is the relay channel capacity [for achievability, don't use relay].

First non-tight example →

3. Zhang '88: Primitive "Stochastically" Degraded Channel (not symmetric)

   Assume $Y \perp\!\!\!\perp Z | X$, $P_{Y|X} = P_{Z|X} \circ Q_{Y|Z}$ for some $Q_{Y|Z}$. Then:

   $$R_0 > \max_{P_X : I(X;Y) = C_{XY}} I(X;Z) - C_{XY} \Rightarrow C(R_0) < C_{XY} + R_0 \,, \text{ where } C_{XY} = \max_{P_X} I(X;Y).$$

   Note: Known that: $R_0 \leq \max_{P_X : I(X;Y) = C_{XY}} I(X;Z) - C_{XY} \Rightarrow C(R_0) = C_{XY} + R_0$. [Intuition: We can send $C_{XY} + R_0$ rate to Z, which will send the R_0 part to decoder.]

   Note: $\max_{P_X : I(X;Y) = C_{XY}} I(X;Z) - C_{XY} < R_0 < \max_{P_X : I(X;Y) = C_{XY}} I(X;Y,Z) - C_{XY} \Rightarrow$ Cutset Bound is $R \leq C_{XY} + R_0$ BUT no equality due to Zhang's result.

Proof: $nR = H(M) = I(M;\hat{M}) + H(M|\hat{M}) \leq I(M;\hat{M}) + n\mathcal{J}_n$, $\mathcal{J}_n \to 0$ as $n \to \infty$

↑ [Fano: $M \to X^n \to (Y^n, Z^n) \to \hat{M}$, $H(M|\hat{M}) \leq H(P_e^{(n)}) + P_e^{(n)} nR = n\Big(\underbrace{\frac{H(P_e^{(n)})}{n} + P_e^{(n)} R}_{\to 0 \text{ as } n \to \infty}\Big)$]

$$\overset{\text{DPI}}{\Rightarrow} nR \leq I(X^n; Y^n, W) + n\mathcal{J}_n$$

↙ Broadcast bound

$$nR \leq I(X^n; Y^n, Z^n) + n\mathcal{J}_n \quad \text{[DPI]}$$
$$\leq \sum_{i=1}^{n} I(X_i; Y_i, Z_i) + n\mathcal{J}_n \quad \text{[memoryless]}$$
$$= nI(X_Q; Y_Q, Z_Q | Q) + n\mathcal{J}_n \quad [Q \sim \text{Unif}\{1,\dots,n\} \perp\!\!\!\perp (X^n, Y^n, Z^n)] \text{ (time sharing)}$$
$$\leq nI(X_Q; Y_Q, Z_Q) + n\mathcal{J}_n \quad [Q \to X_Q \to (Y_Q, Z_Q)]$$
$$\Rightarrow R \leq I(X_Q; Y_Q, Z_Q) + \mathcal{J}_n$$

$I(X_Q; Y_Q, Z_Q | Q)$
$= H(Y_Q, Z_Q | Q)$
$- H(Y_Q, Z_Q | \cancel{Q}, X_Q)$
$\leq I(X_Q; Y_Q, Z_Q)$

MAC bound ↘

$$nR \leq I(X^n; Y^n) + I(W; X^n | Y^n) + n\mathcal{J}_n$$
$$\leq \sum_{i=1}^{n} I(X_i; Y_i) + \underbrace{H(W|Y^n)}_{\leq nR_0} - \underbrace{H(W|X^n, Y^n)}_{\geq 0} + n\mathcal{J}_n \quad \text{[mem'less]}$$
$$\leq \sum_{i=1}^{n} I(X_i; Y_i) + nR_0 + n\mathcal{J}_n$$
$$= nI(X_Q; Y_Q | Q) + nR_0 + n\mathcal{J}_n$$
$$\leq nI(X_Q; Y_Q) + nR_0 + n\mathcal{J}_n$$
$$\Rightarrow R \leq I(X_Q; Y_Q) + R_0 + \mathcal{J}_n$$

Take $n \to \infty$, $P_{X_Q}^{(n)} \to P_X$ WLOG by compactness. $P_{XY} \to I(X;Y)$ cont. in disc. case.

▨

Anuran Makur

③ Improving Cutset Bound: (Wu-Özgür-Xie '16)

From MAC bound:

$$nR \leq nI(X_Q;Y_Q) + H(W|Y^n) - H(W|X^n) + n\mathcal{J}_n \qquad [\text{Symmetry: } W \perp\!\!\!\perp Y^n | X^n]$$

Let $H(W|X^n) = n\varepsilon_n$. First, observe that:

Intuition
1. If $H(W|X^n) \approx 0$, then all Z^n jointly typical with X^n map to W given X^n. Since $P_{Y|X} = P_{Z|X}$, we also expect $H(W|Y^n) \approx 0$. So, $H(W|Y^n) \leq nR_0$ is a loose bound.
2. If $\varepsilon_n \geq \varepsilon$ for large n, then $H(W|X^n) \geq 0$ is a loose bound.

Information inequality: $Y^n \underset{\uparrow\text{same}}{\overset{\text{DMC}}{\longleftrightarrow}} X^n \overset{\text{DMC}}{\longleftrightarrow} Z^n \xrightarrow{f_n} W$, $\varepsilon_n = \dfrac{H(W|X^n)}{n}$. ← Do not need reliable code information.

$\forall n \in \mathbb{N}$, $H(W|Y^n) \leq n\, g(\varepsilon_n)$ for some function $g: \mathbb{R}^+ \to \mathbb{R}^+$ continuous with $g(0)=0$.

↑explicit examples in [Wu-Özgür-Xie '16]

Then, we have: $nR \leq nI(X_Q;Y_Q) + H(W|Y^n) - n\varepsilon_n + n\mathcal{J}_n$

↙ $H(W|Y^n) \leq nR_0$ [card. bound] ↘ $H(W|Y^n) \leq n\,g(\varepsilon_n)$ [info. ineq.]

$$R \leq I(X_Q;Y_Q) + R_0 - \varepsilon_n + \mathcal{J}_n, \qquad R \leq I(X_Q;Y_Q) + g(\varepsilon_n) - \varepsilon_n + \mathcal{J}_n.$$

So, we get:

$$\begin{aligned} R &\leq I(X_Q;Y_Q,Z_Q) + \mathcal{J}_n \\ R &\leq I(X_Q;Y_Q) + R_0 - \varepsilon_n + \mathcal{J}_n \\ R &\leq I(X_Q;Y_Q) + g(\varepsilon_n) - \varepsilon_n + \mathcal{J}_n \end{aligned}, \quad \text{where } 0 \leq \varepsilon_n \leq R_0.$$

By compactness, $\varepsilon_n \to \varepsilon \in [0,R_0]$ because $P_{X_Q}^{(n)} \to P_X$ WLOG as $n \to \infty$. So, letting $n\to\infty$, we derive the following result:

↳ continuity of cond. entropy

Thm: For the symmetric primitive relay channel, if R is achievable, then $\exists P_X$, $\exists \varepsilon \in [0,R_0]$ s.t.:

$$\begin{aligned} R &\leq I(X;Y,Z), \\ R &\leq I(X;Y) + R_0 - \varepsilon, \\ R &\leq I(X;Y) + g(\varepsilon) - \varepsilon. \end{aligned}$$

} better than MAC bound: $R \leq I(X;Y) + R_0$

Remarks:

1. Improvement is not obtained via tensorization.
2. Strictly tighter than cutset bound for $R_0 > 0$. [$\varepsilon > 0 \Rightarrow$ obvious, $\varepsilon = 0 \Rightarrow$ Ineq. (3) is $R \leq I(X;Y)$]
3. As in cutset bound, all 3 inequalities are coupled by P_X.

④ Proof of Information Inequality: (special case)

Suppose $Y^n \overset{\text{DMC}}{\longleftrightarrow} X^n \overset{\text{DMC}}{\longleftrightarrow} Z^n \longrightarrow W = f_n(Z^n) \in \{1,\dots,2^{nR_0}\}$, and $\underline{\frac{1}{n}H(W|X^n) \triangleq \varepsilon_n \xrightarrow{n\to\infty} 0}$.

Thm: $H(W|Y^n) \leq n g(\varepsilon_n)$ for some function $g: \mathbb{R}^+ \to \mathbb{R}^+$ s.t. $\underbrace{0 = g(0) = \lim_{x\to 0^+} g(x)}_{\text{continuous at zero}}$, and for sufficiently large n.

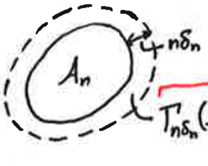
Blowing-Up Lemma: (Ahlswede-Gács-Körner '76, Marton '86)

↳original ↳info. theoretic proof (transportation-cost inequality)

Let $U_1, U_2, \dots, U_n \in \mathcal{U}$ be independent random variables. Suppose $A_n \subseteq \mathcal{U}^n$ satisfies $\mathbb{P}(U^n \in A_n) \geq 2^{-n\varepsilon_n}$ for $\varepsilon_n \xrightarrow{n\to\infty} 0$. Then, $\exists$ sequences $\delta_n \xrightarrow{n\to\infty} 0$ and $\eta_n \xrightarrow{n\to\infty} 0$ such that $\mathbb{P}(U^n \in \Gamma_{n\delta_n}(A_n)) \geq 1 - \eta_n$.

↑only depend on ε_n, not on $A_n \Rightarrow$ can use any A_n with same sequences

Marton's proof: $\eta_n = \dfrac{\sqrt{\varepsilon_n}}{\delta_n}$ and choose δ_n s.t. $\eta_n \xrightarrow{n\to\infty} 0$.
Eg: $\delta_n = \varepsilon_n^{1/4}$, $\eta_n = \varepsilon_n^{1/4}$



$n\delta_n$-blowup of A_n (or) $n\delta_n$-enlargement of A_n

$$\Gamma_{n\delta_n}(A_n) \triangleq \{u^n \in \mathcal{U}^n : \exists v^n \in A_n,\ d_H(u^n,v^n) \leq n\delta_n\}.$$

↑Hamming distance: $d_H(u^n,v^n) = \sum_{i=1}^{n} \mathbb{1}\{u_i \neq v_i\}$

Proof: (Polyanskiy '16)

① Lower Bound $P_{W|X^n}$ using Markov's Inequality:

$$\mathbb{P}\left(P_{W|X^n}(W|X^n) < 2^{-n\sqrt{\varepsilon_n}}\right) = \mathbb{P}\left(\log_2\left(\frac{1}{P_{W|X^n}(W|X^n)}\right) > n\sqrt{\varepsilon_n}\right) \leq \frac{H(W|X^n)}{n\sqrt{\varepsilon_n}} \overset{H(W|X^n) = n\varepsilon_n}{=} \sqrt{\varepsilon_n}$$ [Markov inequality]

Let $S = \{(w, x^n) \in \{1,\dots,2^{nR_0}\} \times \mathcal{X}^n : P_{W|X^n}(w|x^n) \geq 2^{-n\sqrt{\varepsilon_n}}\}$, and note that:

$\mathbb{P}((W, X^n) \in S) \geq 1 - \sqrt{\varepsilon_n}$. [large n]

② Let $f_n^{-1}(\omega) \triangleq \{z^n \in \mathcal{Y}^n : f_n(z^n) = \omega\}$ for any $\omega \in \{1,\dots,2^{nR_0}\}$.
pre-image set

Show $\mathbb{P}(Y^n \in T_{n\delta_n}(f_n^{-1}(W)))$ is large using Blowing-Up Lemma:

For $(w, x^n) \in S$, $P_{W|X^n}(w|x^n) = \mathbb{P}(Z^n \in f_n^{-1}(\omega) | X^n = x^n) \geq 2^{-n\sqrt{\varepsilon_n}}$.

$$\overset{\text{Blow-up}}{\underset{\text{Lemma}}{\Longrightarrow}} \quad \mathbb{P}(Z^n \in T_{n\delta_n}(f_n^{-1}(\omega)) | X^n = x^n) \geq 1 - \eta_n, \quad \text{where } \delta_n = \eta_n = \varepsilon_n^{\frac{1}{8}}.$$

$$\overset{P_{Y|X} \overset{d}{=} P_{Z|X}}{\underset{\text{Symmetry}}{\Longrightarrow}} \quad \mathbb{P}(Y^n \in T_{n\delta_n}(f_n^{-1}(\omega)) | X^n = x^n) \geq 1 - \eta_n$$

$$\mathbb{P}(Y^n \in T_{n\delta_n}(f_n^{-1}(W))) \geq \sum_{(\omega, x^n) \in S} \underbrace{\mathbb{P}(Y^n \in T_{n\delta_n}(f_n^{-1}(\omega)) | X^n = x^n, \cancel{W = \omega})}_{\geq 1 - \eta_n} P_{W,X^n}(\omega, x^n)$$

$Y^n \perp\!\!\!\perp W | X^n$

$$\geq (1 - \eta_n)\, \mathbb{P}((W, X^n) \in S)$$
$$\geq (1 - \eta_n)(1 - \sqrt{\varepsilon_n}) \quad \leftarrow \eta_n = \varepsilon_n^{\frac{1}{8}}$$
$$\geq 1 - 2\varepsilon_n^{\frac{1}{8}} \quad .$$ [large n]

③ Bound $H(W|Y^n)$ using List Decoding + Fano's Inequality:
↳(Ahlswede – Dueck '76)

Let $E = \mathbb{1}\{Y^n \in T_{n\delta_n}(f_n^{-1}(\omega))\}$.
↑indicator r.v.

$$H(W|Y^n) \leq H(W, E|Y^n) = H(E|Y^n) + H(W|E, Y^n)$$
$$\leq \underbrace{H(E)}_{\leq h(2\varepsilon_n^{\frac{1}{8}})} + \underbrace{P_E(0)}_{\leq 2\varepsilon_n^{\frac{1}{8}}} \underbrace{H(W|E=0, Y^n)}_{\leq nR_0} + \underbrace{P_E(1)}_{\approx 1} \underbrace{H(W|E=1, Y^n)}$$

[large n] ← (binary entropy)

If $Y^n \in T_{n\delta_n}(f_n^{-1}(W))$, then
$\exists y^n \in f_n^{-1}(W) \cap \underbrace{\{z^n \in \mathcal{Y}^n : d_H(z^n, y^n) \leq n\delta_n\}}_{\text{Hamming Ball}_{y^n}(n\delta_n)}$.

$$\Rightarrow H(W|Y^n, E=1) \leq \log_2(|\{z^n \in \mathcal{Y}^n : d_H(z^n, y^n) \leq n\delta_n|)$$
$$\leq n(\delta_n \log_2(|\mathcal{Y}| - 1) + h(\delta_n))$$
↑bound on volume of Hamming ball (requires $n\delta_n \geq 1$)

$\delta_n = \varepsilon_n^{\frac{1}{8}}$

$$\Rightarrow H(W|Y^n) \leq h(2\varepsilon_n^{\frac{1}{8}}) + n 2\varepsilon_n^{\frac{1}{8}} R_0 + n(\delta_n \log_2(|\mathcal{Y}| - 1) + h(\delta_n))$$

$$\Rightarrow H(W|Y^n) \leq n\Big(\underbrace{h(2\varepsilon_n^{\frac{1}{8}}) + 2\varepsilon_n^{\frac{1}{8}} R_0 + h(\varepsilon_n^{\frac{1}{8}}) + \varepsilon_n^{\frac{1}{8}} \log_2(|\mathcal{Y}| - 1)}_{\triangleq g(\varepsilon_n) (\longrightarrow 0 \text{ as } \varepsilon_n \to 0)}\Big)$$

$\therefore$ $H(W|Y^n) \leq n g(\varepsilon_n)$.

▨

Remarks:

1. Bound better for larger n.
2. Such bound does not hold without factoring $X^n \to Z^n \to W$. [Eg: $Y_n \leftarrow X^n \rightarrow W$ Let $W = X^n$, $Y^n \perp\!\!\!\perp X^n$. Then, $H(W|X^n) = 0$, $H(W|Y^n) = H(X^n)$.]

// THE END

Anuran Makur
5 Dec 2016

Volume of Hamming Ball:

Consider the set $\mathcal{Y}$ with $2 \leq |\mathcal{Y}| < +\infty$. For $y^n, z^n \in \mathcal{Y}^n$, let $d_H(y^n, z^n) \triangleq \sum_{i=1}^{n} \mathbb{1}\{y_i \neq z_i\}$ denote Hamming distance.
Let $\text{Ball}_{y^n}(nr) \triangleq \{z^n \in \mathcal{Y}^n : d_H(y^n, z^n) \leq nr\}$ denote the Hamming ball of radius nr around $y^n \in \mathcal{Y}^n$.

Prop: For $0 \leq r \leq 1 - \frac{1}{|\mathcal{Y}|}$, and sufficiently large n (s.t. $nr \geq 1$), we have:

$$\frac{1}{n}\log(|\text{Ball}_{y^n}(nr)|) \leq r\log(|\mathcal{Y}|-1) + h(r).$$

any base

[Note: This is asymptotically tight.]

Proof:

$$\begin{aligned}
1 &= (r + (1-r))^n \\
&= \sum_{i=0}^{n} \binom{n}{i} r^i (1-r)^{n-i} \\
&\geq \sum_{i=0}^{\lfloor nr \rfloor} \binom{n}{i} r^i (1-r)^{n-i} \\
&= \sum_{i=0}^{\lfloor nr \rfloor} \binom{n}{i} (|\mathcal{Y}|-1)^i (1-r)^n \left(\frac{r}{(|\mathcal{Y}|-1)(1-r)}\right)^i \\
&\geq (1-r)^n \left(\frac{r}{(|\mathcal{Y}|-1)(1-r)}\right)^{nr} \underbrace{\sum_{i=0}^{\lfloor nr \rfloor} \binom{n}{i} (|\mathcal{Y}|-1)^i}_{= |\text{Ball}_{y^n}(nr)|} \\
&= |\text{Ball}_{y^n}(nr)| \, (1-r)^n \left(\frac{r}{(|\mathcal{Y}|-1)(1-r)}\right)^{nr} = |\text{Ball}_{y^n}(nr)| \, r^{nr} (1-r)^{n(1-r)} (|\mathcal{Y}|-1)^{-nr}
\end{aligned}$$

$$\left[\frac{r}{(|\mathcal{Y}|-1)(1-r)} \leq 1 \Leftrightarrow r \leq |\mathcal{Y}| - 1 - r(|\mathcal{Y}|-1) \Leftrightarrow \underbrace{r|\mathcal{Y}| \leq |\mathcal{Y}|-1}_{\text{this is assumed}}\right]$$

$$\Rightarrow \quad 0 \geq \frac{1}{n}\log(|\text{Ball}_{y^n}(nr)|) + r\log(r) + (1-r)\log(1-r) - r\log(|\mathcal{Y}|-1)$$

$$\Rightarrow \quad \frac{1}{n}\log(|\text{Ball}_{y^n}(nr)|) \leq r\log(|\mathcal{Y}|-1) + \underset{\uparrow \text{ binary entropy}}{h(r)}$$

▨

Intuition:

$$|\text{Ball}_{y^n}(nr)| \leq \underbrace{(|\mathcal{Y}|-1)^{nr}}_{\substack{\text{no. of choices} \\ \text{of replacement} \\ \text{at } nr \text{ locations}}} \cdot \underbrace{\exp(nh(r))}_{\substack{\text{no. of binary strings of length } n \\ \text{with type } (nr\ 1\text{'s},\ n(1-r)\ 0\text{'s})}} \longleftarrow \left[\substack{nr \text{ distance dominates} \\ \text{lower distance strings}}\right]$$

encode changed letters