EE376A Information Theory Lecture 11 - 02/12/2015

Lecture 11

Lecturer: Tsachy Weissman *Scribe: Padmapriya, Jayant, Pinnaree*

1 Joint AEP

Let X, Y be jointly random variables with alphabets $\mathcal{X}$, $\mathcal{Y}$, respectively. Let the source be memoryless so that $(X_i,\, Y_i)$ are i.i.d. $\sim P_{X,Y}$. That is,

$$P(x^n) = \prod_{i=1}^{n} P_X(x_i), \quad P(y^n) = \prod_{i=1}^{n} P_Y(y_i), \quad P(X^n, Y^n) = \prod_{i=1}^{n} P_{X,Y}(x_i, y_i) \tag{1}$$

1.1 Set of Jointly Typical Sequences

Let $A_\epsilon^{(n)}(X,Y)$ denote the set of jointly typical sequences. That is,

$$A_\epsilon^{(n)}(X,Y) = \{(X^n, Y^n) : |-\frac{1}{n}\log P(x^n) - H(X)| < \epsilon, \tag{2}$$

$$|-\frac{1}{n}\log P(y^n) - H(Y)| < \epsilon, \tag{3}$$

$$|-\frac{1}{n}\log P(x^n, y^n) - H(X,Y)| < \epsilon\} \tag{4}$$

Theorem A: If (X^n, Y^n) are formed by i.i.d. $(X_i, Y_i) \sim P_{X,Y}$, then

1.
$$\lim_{n\to\infty} P((X^n, Y^n) \in A_\epsilon^{(n)}(X,Y)) = 1 \tag{5}$$

   By AEP, we have X^n is typical, Y^n is typical, and (X^n, Y^n) is typical too.

2. $\forall \epsilon > 0, \exists n_0 \in \mathbb{N} \ni \forall n > n_0$
$$(1-\epsilon)2^{n[H(X,Y)-\epsilon]} \le |A_\epsilon^{(n)}(X,Y)| \le 2^{n[H(X,Y)+\epsilon]} \tag{6}$$

Theorem B: If $(\tilde{X}^n, \tilde{Y}^n)$ are formed by i.i.d. $(\tilde{X}_i, \tilde{Y}_i) \sim (\tilde{X}, \tilde{Y})$, where $P_{\tilde{X},\tilde{Y}} = P_X \cdot P_Y$ then $\forall \epsilon > 0, \exists n_0 \in \mathbb{N}$ such that $\forall n > n_0$

$$(1-\epsilon)2^{-n[I(X;Y)+3\epsilon]} \le P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)) \le 2^{-n[I(X;Y)-3\epsilon]} \tag{7}$$

Intuition:

$$|A_\epsilon^{(n)}(\tilde{X}, \tilde{Y})| \approx 2^{nH(\tilde{X},\tilde{Y})} \tag{8}$$

$$= 2^{n(H(X)+H(Y))} \tag{9}$$

$$= 2^{nH(X)} \cdot 2^{nH(Y)} \tag{10}$$

$$\approx |A_\epsilon^{(n)}(X)| \cdot |A_\epsilon^{(n)}(Y)| \tag{11}$$

Note that $(\tilde{X}, \tilde{Y})$ are distributed uniformly within a set of size $|A_\epsilon^{(n)}(X)| \cdot |A_\epsilon^{(n)}(Y)|$

$$\Rightarrow P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)) = \frac{|A_\epsilon^{(n)}(X,Y)|}{|A_\epsilon^{(n)}(X)| \cdot |A_\epsilon^{(n)}(Y)|} \tag{12}$$

$$\approx \frac{2^{nH(X,Y)}}{2^{nH(X)} \cdot 2^{nH(Y)}} \tag{13}$$

$$= 2^{-nI(X;Y)} \tag{14}$$

Proof:

$$P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)) = \sum_{(\tilde{x}^n, \tilde{y}^n) \in A_\epsilon^{(n)}(X,Y)} P(\tilde{x}_n) \cdot P(\tilde{y}_n) \tag{15}$$

$$\leq \sum_{(\tilde{x}^n, \tilde{y}^n) \in A_\epsilon^{(n)}(X,Y)} 2^{-nH(X)-\epsilon} \cdot 2^{-nH(Y)-\epsilon} \tag{16}$$

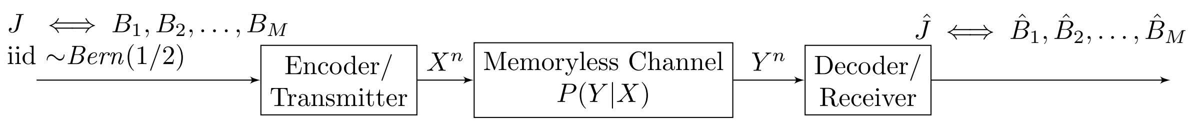
$$= |A_\epsilon^{(n)}(X,Y)| \left[2^{-n(H(X)+H(Y)-2\epsilon} \right] \tag{17}$$

$$\leq 2^{n[H(X,Y)+\epsilon]-[H(X)+H(Y)-2\epsilon]} \tag{18}$$

$$\leq 2^{n[H(X,Y)-[H(X)+H(Y)-3\epsilon]} \tag{19}$$

$$2^{-nI(X;Y)-3\epsilon} \tag{20}$$

2 Communication Setting with Finite Alphabet



J is uniformly distributed on $\{1, 2, \ldots, M\}$. We define our scheme as follows:

Encoder (also known as "codebook"): $\{1, 2, \ldots, M\} \rightarrow X^n$.
That is, codebook $c_n = \{X^n(1), X^n(2), \ldots, X^n(M)\}$

Decoder: $\hat{J}(\cdot) : Y^n \rightarrow \{1, 2, \ldots, M\}$

Rate: Bits per channel use $= \log(m)/n = \log(|c_n|)/n$

3 Direct Theorem

If $R < \max_{P_X} I(X;Y)$, then R is achievable. Equivalently, if $\exists P_X$ s.t. $R < I(X;Y)$, then R is achievable.

Proof:
Fix P_X and a rate $R < I(X;Y)$. Choose $\epsilon = (I(X;Y) - R)/4$. This means that $R < I(X;Y) - 3\epsilon$.
Generate codebook C_n of size $M = \lceil 2^{nR} \rceil$.
$X^n(k)$ are i.i.d. with distribution P_X, $\forall\ k = 1, 2, \ldots, M$ .Then

$$\hat{J}(Y^n) = \begin{cases} j & \text{if} \quad (X^n(j), Y^n) \in A_\epsilon^n(X,Y) \quad \text{and} \quad (X^n(k), Y^n) \notin A_\epsilon^n(X,Y) \forall j \neq k \\ error & \text{otherwise} \end{cases} \tag{21}$$

Denote probability of error using a codebook c^n as $P_e(c^n)$. Thus, $P_e(c^n) = P(\hat{J} \neq J | C_n = c_n)$

$$E[P_e(C_n)] = P(\hat{J} \neq J) \tag{22}$$

$$= \sum_{i=1}^{M} P(\hat{J} \neq J | J = i) P(J = i) \tag{23}$$

$$= P(\hat{J} = J | J = 1) \tag{24}$$

This is because, by symmetry, $P(\hat{J} \neq J | J = i) = P(\hat{J} \neq J | J = j) \quad \forall i, j$.

By union bound, it follows that
$P(\hat{J} \neq J | J = 1)$
$\leq P((X^n(1), Y^n) \notin A_\epsilon^{(n)}(X, Y) + \sum_{k=2}^{M} P((X^n(k), Y^n) \in A_\epsilon^{(n)}(X, Y))$
The first term on the right tends to zero as n tends to infinity. Therefore,

$$P(\hat{J} \neq J | J = 1) \leq \sum_{k=2}^{M} P((X^n(k), Y^n) \in A_\epsilon^{(n)}(X, Y)) \tag{25}$$

$$\leq \sum_{k=2}^{M} P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X, Y)) \tag{26}$$

$$\leq (M - 1) \cdot 2^{-n(I(X;Y) - 3\epsilon)} \tag{27}$$

$$\leq 2^{nR} \cdot 2^{-n(I(X;Y) - 3\epsilon)} \tag{28}$$

$$\leq 2^{-n[I(X;Y) - 3\epsilon - R]} \tag{29}$$

Since $R < I(X; Y) - 3\epsilon$, the expression tends to zero as n tends to infinity.

This means that,

$\exists c_n$ s.t. $|c_n| \geq 2^{nR}$ and $P_e(c_n) \leq E[P_e(C_n)]$
$\Rightarrow \exists c_n$ s.t. $|c_n| \geq 2^{nR}$ and $\lim_{n \to \infty} P_e(c_n) = 0$
$\Rightarrow$ R is achievable.