EE276 Information Theory Lecture 11 - 03/14/2024

Lecture 11: Joint AEP

Lecturer: Tsachy Weissman

1 Joint AEP

We have the following setting:

$$\begin{aligned}
X, Y &\text{ random variables on alphabets } \mathcal{X}, \mathcal{Y} \\
(X,Y) &\sim P_{X,Y} \\
X &\sim P_X \\
Y &\sim P_Y \\
(X_i, Y_i) &\text{ iid } \sim (X,Y) \\
p(x^n) &= \prod_{i=1}^{n} P_X(x_i) \\
p(y^n) &= \prod_{i=1}^{n} P_Y(y_i) \\
p(x^n, y^n) &= \prod_{i=1}^{n} P_{X,Y}(x_i, y_i)
\end{aligned}$$

Definition 1. *The set of jointly ϵ-typical sequences is:*

$$\begin{aligned}
A_\epsilon^{(n)}(X,Y) = \Bigg\{ (x^n, y^n) : &\left| -\frac{1}{n}\log p(x^n) - H(X) \right| \leq \epsilon, \\
&\left| -\frac{1}{n}\log p(y^n) - H(Y) \right| \leq \epsilon, \\
&\left| -\frac{1}{n}\log p(x^n, y^n) - H(X,Y) \right| \leq \epsilon, \Bigg\}
\end{aligned}$$

Theorem 2. ***Joint AEP.***

Part A. *If (X^n, Y^n) formed by iid (X_i, Y_i):*

1. $P\left((X^n, Y^n) \in A_\epsilon^{(n)}(X,Y) \right) \xrightarrow{n\to\infty} 1$
2. $(1-\epsilon)2^{n(H(X,Y)-\epsilon)} \leq \left| A_\epsilon^{(n)}(X,Y) \right| \leq 2^{n(H(X,Y)+\epsilon)}$*, where the first inequality holds for sufficiently large n, and the second inequality holds for all n.*

Proof
We apply AEP, and convergence in probability on the three conditions of the jointly typical set. That is, there exists n_1, n_2, n_3 such that for all $n > n_1$, we have

$$\Pr\left\{ \left| -\frac{1}{n}\log p(x^n) - H(X) \right| \geq \epsilon \right\} < \epsilon/3,$$

and for all $n > n_2$, we have

$$\Pr\left\{\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| \geq \epsilon\right\} < \epsilon/3,$$

and for all $n > n_3$, we have

$$\Pr\left\{\left|-\frac{1}{n}\log p(x^n, y^n) - H(X, Y)\right| \geq \epsilon\right\} < \epsilon/3.$$

All three apply for n greater than the largest of n_1, n_2, n_3. Therefore the probability of the union the set of (x^n, y^n) satisfying these inequalities must be less than ϵ, and for n sufficiently large, the probability of the set $A_\epsilon^{(n)}$ is greater than $1 - \epsilon$.

Upper Bound:

$$\begin{aligned}
1 &= \sum p(x^n, y^n) \\
&\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}(X,Y)} p(x^n, y^n) \\
&\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}(X,Y)} 2^{-n(H(X,Y)+\epsilon)}, \text{ by definition of typicality} \\
&= 2^{-n(H(X,Y)+\epsilon)} \left|A_\epsilon^{(n)}(X, Y)\right| \\
&\Rightarrow \left|A_\epsilon^{(n)}(X, Y)\right| \leq 2^{n(H(X,Y)+\epsilon)}
\end{aligned}$$

Lower Bound:
By Part 1, $P\left((X^n, Y^n) \in A_\epsilon^{(n)}(X, Y)\right) \xrightarrow{n\to\infty} 1$.
Thus, for large n:

$$\begin{aligned}
1 - \epsilon &\leq P((X^n, Y^n) \in A_\epsilon^{(n)}(X, Y)) \\
&\leq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} 2^{-n(H(X,Y)-\epsilon)} \\
&= 2^{-n(H(X,Y)-\epsilon)} \left|A_\epsilon^{(n)}(X, Y)\right| \\
&\Rightarrow \left|A_\epsilon^{(n)}(X, Y)\right| \geq (1 - \epsilon)2^{n(H(X,Y)-\epsilon)}
\end{aligned}$$

□

Part B. For $(\widetilde{X}^n, \widetilde{Y}^n)$ where $P_{\widetilde{X},\widetilde{Y}} = P_X \times P_Y$ (essentially you have sequences X^n, Y^n which are drawn from P_X and P_Y **independently**):

$$(1 - \epsilon)2^{-n(I(X;Y)+3\epsilon)} \leq P\left\{(\widetilde{X}^n, \widetilde{Y}^n) \in A_\epsilon^{(n)}(X, Y)\right\} \leq 2^{-n(I(X;Y)-3\epsilon)}$$

2 Recap: Communication problem

Recall the communication problem

$$J \sim \text{uniform} \in \{1,2,...,M\} \to \boxed{\text{encoder}} \xrightarrow{X^n} \boxed{\text{memoryless channel } P_{Y|X}} \xrightarrow{Y^n} \boxed{\text{decoder}} \to \hat{J}$$

- rate $= \frac{\log M}{n}$ $(\frac{\text{bits}}{\text{channel use}})$
- probability of error $P_e = P(\hat{J} \neq J)$
- main result: the maximum rate of reliable communication $C = \max_{P_X} I(X;Y)$

We can interpret the main result as two parts:

- Direct part: if $R < \max_{P_X} I(X;Y)$, then R is achievable, i.e., $\exists$ schemes with rate $\geq R$ and $P_e \xrightarrow{n\to\infty} 0$.
- Converse part: if $R > \max_{P_X} I(X;Y)$, then R is <u>not</u> achievable.

We are going to prove the direct part in this lecture.

3 Joint AEP

Suppose $(X,Y) \sim P_{X,Y}$, X,Y takes values from finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Therefore, (X,Y) has alphabet $\mathcal{X} \times \mathcal{Y}$, where '$\times$' denotes direct product. In this setting, the set of of jointly ϵ-typical sequences is:

$$\begin{aligned} A_\epsilon^{(n)}(X,Y) = \Big\{(x^n, y^n) :& |-\frac{1}{n}\log P(x^n) - H(X)| \leq \epsilon, \\ & |-\frac{1}{n}\log P(y^n) - H(Y)| \leq \epsilon, \\ & |-\frac{1}{n}\log P(x^n, y^n) - H(X,Y)| \leq \epsilon\Big\} \end{aligned} \tag{1}$$

An illustration of the joint AEP $A_\epsilon^n(X,Y)$ is shown in Fig. 1.

Part A

Recall that:
Theorem: If (X_i, Y_i) iid $\sim (X,Y)$, then for $\forall \epsilon > 0$,

1. $P\Big((X^n, Y^n) \in A_\epsilon^{(n)}(X,Y)\Big) \xrightarrow{n\to\infty} 1.$
2. $(1-\epsilon)2^{n(H(X,Y)-\epsilon)} \leq |A_\epsilon^{(n)}(X,Y)| \leq 2^{(nH(X,Y)+\epsilon)}$, for all sufficiently large n.

Part B

Suppose now: $\tilde{X}^n \overset{d}{=} X^n$, $\tilde{Y}^n \overset{d}{=} Y^n$, and $\tilde{X}^n, \tilde{Y}^n$ are <u>independent</u>, where '$\overset{d}{=}$' means equality in distribution. Then:

1. $\tilde{X}^n \approx$ uniformly distributed on $A_n^\epsilon(X)$
2. $\tilde{Y}^n \approx$ uniformly distributed on $A_n^\epsilon(Y)$
3. $\tilde{X}^n, \tilde{Y}^n$ are <u>independent</u> $\Rightarrow$ $(\tilde{X}^n, \tilde{Y}^n) \approx$ uniformly distributed on $A_n^\epsilon(X) \times A_n^\epsilon(Y)$.

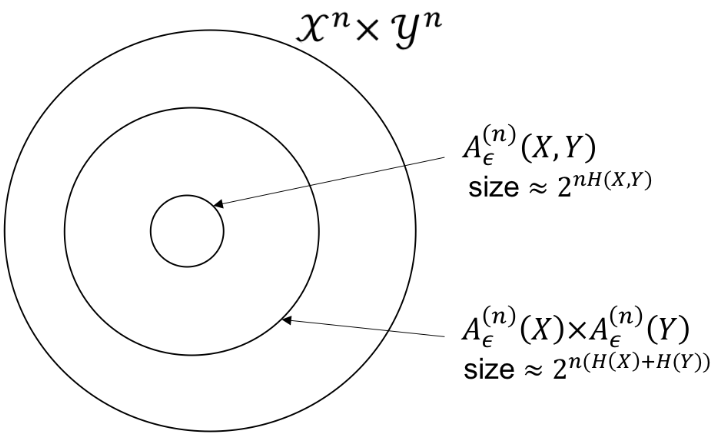
$\mathcal{X}^n \times \mathcal{Y}^n$
$A_\epsilon^{(n)}(X,Y)$
size ≈ $2^{nH(X,Y)}$
$A_\epsilon^{(n)}(X) \times A_\epsilon^{(n)}(Y)$
size ≈ $2^{n(H(X)+H(Y))}$

Figure 1: Illustration of the joint AEP set.

With the above properties, we arrive at

$$P\Big((\tilde{X^n}, \tilde{Y^n}) \in A_\epsilon^{(n)}(X,Y)\Big) \approx \frac{|A_\epsilon^{(n)}(X,Y)|}{|A_n^\epsilon(X) \times A_n^\epsilon(Y)|} \approx \frac{2^{nH(X,Y)}}{2^{nH(X)}2^{nH(Y)}} = 2^{-nI(X;Y)} \tag{2}$$

More rigorously, we have

Theorem: For $\forall \epsilon > 0$ and sufficiently large n, the probability $(\tilde{X}^n, \tilde{Y}^n)$, where $\tilde{X}^n$, $\tilde{Y}^n$ are drawn from P_X and P_Y independently, falls into the jointly typical set satisfies

$$(1-\epsilon) \cdot 2^{-n(I(X;Y)+3\epsilon)} \leq P\Big((\tilde{X^n}, \tilde{Y^n}) \in A_\epsilon^{(n)}(X,Y)\Big) \leq 2^{-n(I(X;Y)-3\epsilon)} \tag{3}$$

This states that in the case of a pair of sequences, it is very unlikely for a pair of independent sequences to look as if they came from a joint source described by $P(X,Y)$ with the exponent in the probability being $-nI(X;Y)$.

Proof:

By definition

$$P\Big((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)\Big) = \sum_{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)} P(\tilde{X}^n, \tilde{Y}^n) \tag{4}$$

As previously shown, we have

$$(1-\epsilon)2^{n(H(X,Y)-\epsilon)} \leq |A_\epsilon^{(n)}(X,Y)| \leq 2^{n(H(X,Y)+\epsilon)} \tag{5}$$

Also, since $\tilde{X}$ and $\tilde{Y}$ satisfies (by the typicality of each of them)

$$2^{-n(H(X)+\epsilon)} \leq P(\tilde{X}^n) \leq 2^{-n(H(X)-\epsilon)} \tag{6}$$

$$2^{-n(H(Y)+\epsilon)} \leq P(\tilde{Y}^n) \leq 2^{-n(H(Y)-\epsilon)} \tag{7}$$

Since $\tilde{X}$ and $\tilde{Y}$ are independent, by Inequalities 6, 7 we have

$$2^{-n(H(X)+H(Y)+2\epsilon)} \leq P(\tilde{X}^n, \tilde{Y}^n) = P(\tilde{X}^n)P(\tilde{Y}^n) \leq 2^{-n(H(X)+H(Y)-2\epsilon)} \tag{8}$$

Thus by relations 4, 8

$$\sum_{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)} 2^{-n(H(X)+H(Y)+2\epsilon)} \leq P\Big((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)\Big) \leq \sum_{(\tilde{X^n}, \tilde{Y^n}) \in A_\epsilon^{(n)}(X,Y)} 2^{-n(H(X)+H(Y)-2\epsilon)} \tag{9}$$

By Inequality 5

$$(1-\epsilon)\cdot 2^{n(H(X,Y)-\epsilon)}2^{-n(H(X)+H(Y)+2\epsilon)} \leq P\Big((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)\Big) \leq 2^{n(H(X,Y)+\epsilon)}2^{-n(H(X)+H(Y)-2\epsilon)} \tag{10}$$

$$(1-\epsilon)\cdot 2^{-n(I(X;Y)+3\epsilon)} \leq P\Big((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)\Big) \leq 2^{-n(I(X;Y)-3\epsilon)} \tag{11}$$

4 Direct Theorem

Theorem

If $R < \max_{P_X} I(X;Y)$, then R is achievable (i.e., $\exists$ schemes with rate $\geq R$ and $P_e \xrightarrow{n\to\infty} 0$).

Rough idea in Proof
To establish a scheme, we are given a rate $R < I(X;Y)$. We need to show the existence of an achievable scheme of codebook and decoding rule with rate R. As illustrated in Figure 2, we can randomly selected 2^{nR} code words from the $A_\epsilon^{(n)}(X)$ as our codebook c_n. By AEP on one variable, we know

$$P(X^n \in A_\epsilon^{(n)}(X)) \approx 1 \tag{12}$$

then we can just select x^n i.i.d. from $P(X^n)$ to construct our codebook c_n.

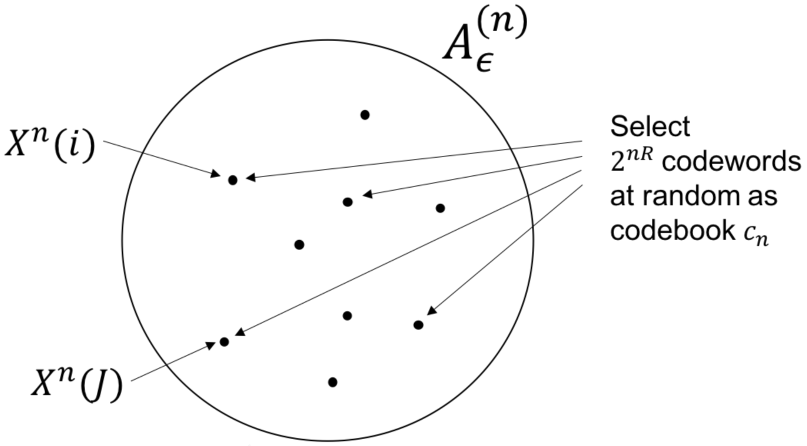

Figure 2: Illustration of selecting the codebook. In this case, the original message is J and i is another message other than J.

Suppose we wish to send a message J (also in Figure 2), and the signals sent is $X^n(J)$ and channel output is Y^n. Then we have

$$P(Y^n \text{ is jointly typical with } X^n(J)) \approx 1 \tag{13}$$

and for a particular $i \neq J$

$$P(Y^n \text{ is jointly typical with } X^n(i)) \approx 2^{-nI(X;Y)} \tag{14}$$

using the indepndence of Y^n and $X^n(i)$. Hence by union bound (provided $R < I(X;Y)$)

$$P(Y^n \text{ is jointly typical with } X^n(i) \text{ for any } i \neq J) \leq 2^{-n(I(X;Y)-R)} \text{ (very small)} \tag{15}$$

We can conclude that joint typicality decoding will be reliable for $R < I(X;Y)$.

Proof:
For a fixed probability distribution P_X, and rate $R < I(X;Y)$, we need to prove that R is reliable.

Let's take a sufficiently small $\epsilon > 0$ that makes the rate satisfy $R < I(X;Y) - 3\epsilon$. Generate a codebook, c_n, with size $M = \lceil 2^{nR} \rceil$ randomly by generating independent sequences $X^n(1), X^n(2), ..., X^n(M)$ where each of them is iid $\sim P_X$. The decoder is the *joint typicality decoder*:

$$\hat{J} = \hat{J}(Y^n) = \begin{cases} j, & \text{if } (X^n(j), Y^n) \in A_\epsilon^{(n)}(X,Y) \text{ and } (X^n(k), Y^n) \notin A_\epsilon^{(n)}(X,Y), \forall k \neq j \\ error, & \text{otherwise} \end{cases} \tag{16}$$

In the situation of correctly decoding, the received symbol is jointly typical with the sent symbol and not jointly typical with any other symbols. Otherwise, the decoder makes an error either because it cannot find such a symbol in the codebook or because it finds more than one. Denoting the probability of error conditioned by a specific codebook as

$$P_e(c_n) = P(\hat{J} \neq J | C_n = c_n) \tag{17}$$

Let's prove its expectation vanishes as n approaches infinity.

$$\begin{aligned}
E[P_e(c_n)] &= P(\hat{J} \neq J) = \Sigma_{j=1}^M P(\hat{J} \neq J | J = j) P(J = j) = P(\hat{J} | J = 1) && \text{(18a)} \\
&\leq P((X^n(1), Y^n) \notin A_\epsilon^{(n)}(X,Y) | J = 1) + \Sigma_{j=2}^M P((X^n(j), Y^n) \in A_\epsilon^n(X,Y) | J = 1) && \text{(18b)} \\
&= P((X^n, Y^n) \notin A_\epsilon^n(X,Y)) + (M-1) P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)) && \text{(18c)} \\
&\leq 2^{nR} \cdot 2^{-n(I(X;Y) - 3\epsilon)} && \text{(18d)} \\
&= 2^{-n(I(X;Y) - 3\epsilon - R)} && \text{(18e)} \\
&\xrightarrow{n \to \infty} 0 && \text{(18f)}
\end{aligned}$$

Inequality (18a) applies the Law of total probability and that $P(\hat{J} \neq J | J = j) = P(\hat{J} \neq J | J = i)$ by symmetry of the scheme. Inequality (18b) applies the union bound. In inequality (18c), the first term converges to 0 as $n \to \infty$, and the second term applies the joint AEP conclusion part B. According to the assumption that $R < I(X;Y) - 3\epsilon$, expression (18e) converges to 0 as $n \to \infty$.

Note 1: $\exists c_n$, s.t. $|c_n| \geq 2^{nR}$ and $P_e(c_n) \leq E[P_e(c_n)]$. This implies

(1) $\exists$ a sequence of $\{c_n\}_{n>=1}$ with $|c_n| \geq 2^{nR}$ and $P_e \xrightarrow{n \to \infty} 0$

(2) R is achievable

Note 2: Our notion of reliability is $P_e = P(\hat{J} \neq J) = \Sigma_{j=1}^M P(\hat{J} \neq J | J = j) P(J = j)$, which is the average probability of error over all symbols. However, one can consider a more stringent criterion $P_{max} = \max_{1 \leq j \leq M} P(\hat{J} \neq J | J = j)$. The exercise below shows that the direct part holds even for this criterion.

Exercise
Show that given c_n with $P_e(c_n)$, $\exists c_n'$ s.t. $|c_n'| \geq \frac{1}{2}|c_n|$ and $P_{max}(c_n') \leq 2P_e(c_n)$.

Proof:
We prove this using an expurgation argument. We remove the $|c_n|/2$ codewords with largest P_e and let c_n' be

the set of the remaining codewords. Clearly, $|c'_n| \geq \frac{1}{2}|c_n|$ is satisfied. We wish to show $P_{max}(c'_n) \leq 2P_e(c_n)$. It can be proved by contradiction.
Assume $P_{max}(c'_n) > 2P_e(c_n)$, i.e., the largest P_e in the smaller half in c_n would be larger than $P_e(c_n)$, which is the average of all the P_e's in c_n. Since the error probabilities in $c_n \setminus c'_n$ are at least $P_{max}(c'_n)$, the average $P_e(c_n) > \frac{1}{2}P_{max}(c'_n)$. But since $P_{max}(c'_n) > 2P_e(c_n)$, we get $P_e(c_n) > P_e(c_n)$, a contradiction.
To show that the rate is unchanged, the rate with c'_n is

$$R' \geq \frac{\log(\frac{1}{2}2^{nR})}{n} = R - \frac{1}{n}. \tag{19}$$

Hence as $n \to \infty$, $R' \to R$ is unchanged.