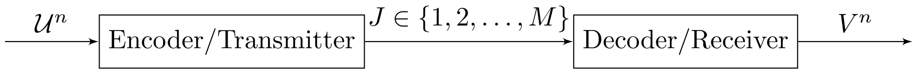
EE376A Information Theory Lecture 17 - 03/05/2015

Lecture 17

Lecturer: Tsachy Weissman *Scribe: Bai Jiang, Rui Fu*

1 Recap

This section is part of lecture 16 scribed by Michael Celentano, Zackary Stratton, Nick White. Many thanks to them.



A scheme is characterized by:

- n, M
- An encoder, mapping from $\mathcal{U}^n$ to $\{1, 2, ..., M\}$ ($\log M$ bits used to encode a symbol sequence, where a symbol sequence is U^n and a symbol is U_i)
- A decoder, mapping from $\{1, 2, ..., M\}$ to $\mathcal{V}^n$

In working with lossy compression, we examine two things:

1. rate $= \frac{\log M}{N} \frac{\text{bits}}{\text{symbol}}$
2. distortion $= d(U^n, V^n) = \frac{1}{n}\sum_{i=1}^{n} d(U_i, V_i)$ (we always specify distortion on a per-symbol basis, and then average the distortions to arrive at $d(U^n, V^n)$)

There's a trade-off between rate and $\frac{\text{distortion}}{\text{symbol}}$. Distortion theory deals with this trade-off.

Definition 1. *(R,D) is <u>achievable</u> if* $\forall \epsilon > 0 \exists$ *scheme (n,M,encoder,decoder) such that* $\frac{\log M}{n} \leq R + \epsilon$ *and* $\mathrm{E}[d(U^n, V^n)] \leq D + \epsilon$

Definition 2. $R(D) \triangleq \inf\{R' : (R', D)$ *is achievable*$\}$

Definition 3. $R^{(I)}(D) \triangleq \min_{\mathrm{E}[d(U,V)] \leq D} I(U; V)$

Theorem 4. $R(D) = R^{(I)}(D)$

2 Proof of Direct Part $R(D) \leq R^{(I)}(D)$

Theorem 4 consists of the direct part $R(D) \leq R^{(I)}(D)$ and the converse part $R(D) \geq R^{(I)}(D)$. This lecture will deal with the former.

2.1 An Equivalent Statement

First, we show the equivalence of the following statements

$$\begin{aligned}R(D) \le R^{(I)}(D) &\Longleftrightarrow R(D) \le \min\{I(U;V) : U, V \text{ s.t. } \mathrm{E}[d(U,V)] \le D\}\\ &\Longleftrightarrow \text{If } U, V \text{ s.t. } \mathrm{E}[d(U,V)] \le D, \text{ then } R(D) \le I(U;V)\\ &\Longleftrightarrow \text{If } U, V \text{ s.t. } \mathrm{E}[d(U,V)] \le D, \text{ then } (R,D) \text{ is achievable for any } R > I(U;V).\end{aligned}$$

Proof The first and second lines follow the definition of $R^{(I)}(D)$. For the last line, it only suffices to show

$$R(D) \le I(U;V) \Longleftrightarrow (R,D) \text{ is achievable for any } R > I(U;V).$$

- For the ⇒ part, consider any $R > I(U;V)$,

  $$R > I(U;V) \ge R(D) = \inf\{R' : (R',D) \text{ is achievable}\}$$

  thus (R,D) is achievable.

- For the ⇐ part, consider some $R'' = I(U;V) + \epsilon$. The assumption (R,D) is achievable for any $R > I(U;V)$ implies that (R'',D) is achievable, and thereafter

  $$R(D) = \inf\{R' : (R',D) \text{ is achievable}\} \le R'' = I(U;V) + \epsilon.$$

  Letting ϵ be arbitrarily small yields $R(D) \le I(U;V)$.

□

In the remaining of the lecture, we are going to prove the equivalent statement instead of $R(D) \le R^{(I)}(D)$. That's to show

$$(R,D) \text{ is the achievable for any } U, V \text{ s.t. } \mathrm{E}[d(U,V)] \le D \text{ and any } R > I(U;V).$$

2.2 Two Useful Lemmas

The proof uses two lemmas appearing as Problem 2 in HW 7 and Problem 7 in HW7, respectively. Let's recall them in advance.

Lemma 5. ***(Joint Typicality Lemma)*** *Suppose* $u^n \in \mathcal{T}_{\delta'}(U)$, $0 < \delta' < \delta$ *and* V_i*'s* $\overset{i.i.d.}{\sim} V$,

$$2^{-n(I(U;V)+\epsilon(\delta))} \le \mathbb{P}\left((u^n, V^n) \in \mathcal{T}_\delta(U,V)\right)$$

for sufficiently large n *and some* $\epsilon(\delta) > 0$ *where* $\lim_{\delta\to 0}\epsilon(\delta) = 0$.

Lemma 6. ***(Typical Average Lemma)***

$$(u^n, v^n) \in \mathcal{T}_\delta(U,V) \Longrightarrow d(u^n, v^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} d(u_i, v_i) \le (1+\delta)\,\mathrm{E}[d(U,V)]$$

2.3 Proof of the Equivalent Statement

For fixed U, V s.t. $\mathrm{E}[d(U,V)] \leq D$ and $R > I(U;V)$, our goal is to show the achieveability of (R, D).
Proof Take $M = \lfloor 2^{nR} \rfloor$. Denote by $C_n = \{V^n(1), V^n(2), ..., V^n(M)\}$ the random codebook which is generated by V_i's $\overset{i.i.d.}{\sim} V$ and independent of U. Let $d(u^n, C_n) = \min_{V^n \in C_n} d(u^n, V^n)$.
For sufficiently small $0 < \delta' < \delta$ which appear in Lemma 1, the assumption $R > I(U, V)$ implies $R > I(U;V) + \epsilon(\delta)$. For any $u^n \in \mathcal{T}_{\delta'}(U)$ and sufficiently large n,

$$\begin{aligned}
\mathbb{P}\left(d(u^n, C_n) > D(1+\delta)\right) &= \mathbb{P}\left(d(u^n, V^n(i)) > D(1+\delta) \text{ for } i = 1, 2, ..., M\right) && \text{(Definition of } d(u^n, C_n))\\
&= \mathbb{P}\left(d(u^n, V^n(1)) > D(1+\delta)\right)^M && (V_i \overset{i.i.d.}{\sim} V)\\
&\leq \mathbb{P}\left(d(u^n, V^n(1)) > \mathrm{E}[d(U,V)](1+\delta)\right)^M && \text{(Assumption of } \mathrm{E}[d(U,V)] \leq D)\\
&\leq \mathbb{P}\left((u^n, V^n(1)) \notin \mathcal{T}_\delta(U,V)\right)^M && \text{(Inverse-negative of Lemma 2)}\\
&= \left[1 - \mathbb{P}\left((u^n, V^n(1)) \in \mathcal{T}_\delta(U,V)\right)\right]^M\\
&\leq \left[1 - 2^{-n(I(U;V)+\epsilon(\delta))}\right]^M && \text{(Lemma 1 with } u^n \in \mathcal{T}_{\delta'}(U) \text{ and large } n)\\
&\leq \exp\left(-M \cdot 2^{-n(I(U;V)+\epsilon(\delta))}\right) && (1 - x \leq e^{-x})
\end{aligned}$$

So far, we have an upper bound of $\mathbb{P}\left(d(u^n, C_n) > D(1+\delta)\right)$ for any $u^n \in \mathcal{T}_{\delta'}(U)$ and sufficiently large n.

$$\mathbb{P}\left(d(u^n, C_n) > D(1+\delta)\right) \leq \exp\left(-M \cdot 2^{-n(I(U;V)+\epsilon(\delta))}\right) \tag{1}$$

Then for $U_i \overset{i.i.d.}{\sim} U$,

$$\begin{aligned}
\mathbb{P}\left(d(U^n, C_n) > D(1+\delta)\right) &= \sum_{u^n \in \mathcal{T}_{\delta'}(U)} \mathbb{P}\left(d(u^n, C_n) > D(1+\delta), U^n = u^n\right)\\
&\quad + \sum_{u^n \notin \mathcal{T}_{\delta'}(U)} \mathbb{P}\left(d(u^n, C_n) > D(1+\delta), U^n = u^n\right)\\
&\leq \sum_{u^n \in \mathcal{T}_{\delta'}(U)} \mathbb{P}\left(d(u^n, C_n) > D(1+\delta)\right) \mathbb{P}\left(U^n = u^n\right) && (U^n \text{ independent of } C_n)\\
&\quad + \mathbb{P}\left(U^n \notin \mathcal{T}_{\delta'}(U)\right)\\
&\leq \exp\left(-M \cdot 2^{-n(I(U;V)+\epsilon(\delta))}\right) + \mathbb{P}\left(U^n \notin \mathcal{T}_{\delta'}(U)\right) && \text{(Upper bound in (1))}
\end{aligned}$$

where the first term goes to 0 as $n \to \infty$ because

$$M = \lfloor 2^{nR} \rfloor, \quad R > I(U;V) + \epsilon(\delta),$$

and the second term goes to 0 as $n \to \infty$ because of AEP. Thus for $U_i \overset{i.i.d.}{\sim} U$,

$$\mathbb{P}\left(d(U^n, C_n) > D(1+\delta)\right) \to 0 \text{ as } n \to \infty \tag{2}$$

Further, let $d(C_n) = \mathrm{E}\left(d(U^n, C_n) | C_n\right)$ be the average distortion by random codebook C_n, and thus $d(c_n) = \mathrm{E}\left(d(U^n, c_n) | C_n = c_n\right) = \mathrm{E}\left(d(U^n, c_n)\right)$ (C_n is independent of U^n) is the average distortion by a realization c_n of C_n.

$$\begin{aligned}
\mathrm{E}[d(C_n)] &= \mathrm{E}\left[\mathrm{E}\left(d(U^n, C_n) | C_n\right)\right]\\
&= \mathrm{E}\left(d(U^n, C_n)\right) && \text{(tower property)}\\
&\leq \mathbb{P}\left(d(U^n, C^n) > D(1+\delta)\right) D_{max} && (D_{max} \triangleq \max_{u \in \mathcal{U}, v \in \mathcal{V}} d(u,v))\\
&\quad + \mathbb{P}\left(d(U^n, C^n) \leq D(1+\delta)\right) D(1+\delta)\\
&\to D(1+\delta) \text{ as } n \to \infty && \text{(Limiting result in (2))}
\end{aligned}$$

It implies that

$$\mathrm{E}[d(C_n)] < D + 2\delta D_{max} \text{ for sufficiently large } n,$$

which further implies existence of c_n, a realization of C_n, satisfying

$$d(c_n) \leq \mathrm{E}[d(C_n)] < D + 2\delta D_{max} \text{ for sufficiently large } n.$$

Taking arbitrarily small δ and sufficiently large n, we can get the average distortion $d(c_n)$ arbitrarily close to D. And the size of codebook

$$|c_n| = M = \lfloor 2^{nR} \rfloor \leq 2^{nR}.$$

(R, D) is achieved by the scheme consisting of n,M,the encoder $u^n \in \mathcal{T}_{\delta'}(U) \mapsto J \in \{1, ..., M\}$, and the decoder $J \in \{1, ..., M\} \mapsto v^n(J) \in c_n$. $\square$