Multi-use Information Theory 2 october 25th, 2010

Lecture 3

Lecturer: Haim Permuter *Scribe: Iddo Naiss*

I. Rate distortion

As mentioned in the previous class, we have both the operational and the mathematical definition of the rate distortion. The operational definition is that given a distortion D, $R(D)$ is the infimum of all achievable rates, where a rate R is achievable if there exists a sequence of codes such that the limit of the expected distortion is lesser than the distortion D. The mathematical definition is simply

$$R^{(I)}(D) = \min_{p(\hat{x}|x):E[d(\hat{X},X)]\leq D} I(X;\hat{X}). \tag{1}$$

Theorem 1 The operational definition $R(D)$ is equal to $R^{(I)}(D)$.

Proof: The proof is divided to two parts. First, the converse: let R be an achievable rate with distortion D. Let us fix a code $(2^{nR}, n)$ with distortion D, i.e., $\frac{1}{n}E[d(X^n, \hat{X}^n(T))] \leq D$. Then

$$\begin{aligned}
nR &\geq H(T) \\
&\overset{(a)}{\geq} I(X^n;T) \\
&\overset{(b)}{\geq} I(X^n;T,\hat{X}^n) \\
&\geq I(X^n;\hat{X}^n) \\
&= H(X^n) - H(X^n|\hat{X}^n) \\
&= \sum_{i=1}^{n} \Big(H(X_i|X^{i-1}) - H(X_i|X^{i-1},\hat{X}_n) \Big) \\
&\overset{(c)}{\geq} \sum_{i=1}^{n} \Big(H(X_i) - H(X_i|\hat{X}_i) \Big) \\
&= \sum_{i=1}^{n} I(X_i;\hat{X}_i),
\end{aligned} \tag{2}$$

where

(a) follows from subtracting $H(T|X^n)$ or from the data processing inequality,

(b) follow the fact that $\hat{X}^n$ is a function of T which implies that $H(X^n|T) = H(X^n|T,\hat{X}^n)$.

(c) Follows from the fact that X^n is i.i.d and the fact that conditioning reduces the entropy.

Now, let Q be a r.v. with the uniform distribution over $\{1, 2, ..., n\}$, and independent of X^n. Lets denote X_Q the element Q^{th} of the sequence X^n. Now consider:

$$\begin{aligned}
R &\geq \frac{1}{n}\sum_{i=1}^{n} I(X_i; \hat{X}_i) \\
&= I(X_Q; \hat{X}_Q|Q) \\
&\overset{(a)}{=} I(X_Q; \hat{X}_Q, Q) \\
&\overset{(b)}{\geq} I(X_Q; \hat{X}_Q) \\
&\overset{(c)}{=} I(X, \hat{X}),
\end{aligned} \tag{3}$$

where

(a) follows from the fact that for all $q \in Q$, $p_{X_Q|Q} = p_{X|Q} = p_X = p_{X_Q}$, which implies that X_Q is independent of Q. In other words X_i is i.i.d and therefore X_Q is independent of Q.

(b) follows from the fact that conditioning reduces entropy, and hence $H(X_Q|\hat{X}_Q, Q) \leq H(X_Q|\hat{X}_Q)$

(c) follows from simply letting $X_Q = X$ and $\hat{X}_Q = \hat{X}$.

As for the distortion,

$$\begin{aligned}
E[d(X, \hat{X})] &= E[d(X_Q, \hat{X}_Q)] \\
&\overset{(a)}{=} E_Q\left[E[d(X_Q, \hat{X}_Q)|Q]\right] \\
&= \frac{1}{n}\sum_{q=1}^{n} E[d(X_q, \hat{X}_q)] \\
&= \frac{1}{n}E[d(X^n, \hat{X}^n)] \\
&\overset{(b)}{\leq} D,
\end{aligned} \tag{4}$$

where (a) is the smoothing property of the expectation, and (b) is due to the code construction. Therefore, we have $R(D) \geq R^{(I)}(D)$. ■

The next step is to show achievability of $R^{(I)}(D)$, and the theorem is proven.