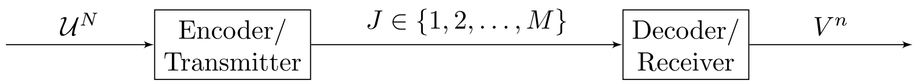
EE376A Information Theory Lecture 1 - 03/03/2015

Lecture 16: Lossy Compression & Rate Distortion Theory

Lecturer: Tsachy Weissman *Scribe: Michael Celentano, Zackary Stratton, Nick White*

1 Definitions



A scheme is characterized by:

- N, M
- An encoder, mapping from $\mathcal{U}^N$ to $\{1, 2, ..., M\}$ (logM 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 $\mathbb{E}[d(U^N, V^N)] \leq D + \epsilon$*

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

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

Theorem 4. $R(D) = R^{(I)}(D)$
NOTE *that $R(D)$ is something we can't solve for (solution space is too large!), but $R^{(I)}(D)$ is something we can solve for (solution space is reasonable).*

Theorem 5. *$R(D)$ is convex,*
i.e. $\forall 0 < \alpha < 1$, D_0 , D_1 : $R(\alpha D_0 + (1-\alpha)D_1) \leq \alpha R(D_0) + (1-\alpha)R(D_1)$

Sketch of proof: We consider a "time-sharing" scheme for encoding N bits. We encode the first αN bits using a "good" scheme for distortion $D = D_0$ and encode the last $(1-\alpha)N$ bits using a "good" scheme for $D = D_1$. Overall, the number of bits in the compressed message is $N\alpha R(D_0) + N(1-\alpha)R(D_1)$, so that the rate is $\alpha R(D_0) + (1-\alpha)R(D_1)$. Further, the expected distortion is the average, weighted by α between the distortions between the two different schemes, i.e. $\alpha D_0 + (1-\alpha)D_1$. We therefore have constructed a scheme which achieves distortion $\alpha D_0 + (1-\alpha)D_1$ with rate $\alpha R(D_0) + (1-\alpha)R(D_1)$, and the optimal scheme can only do better. That is

$$R(\alpha D_0 + (1-\alpha)D_1) \leq \alpha R(D_0) + (1-\alpha)R(D_1)$$

as desired.

2 Examples

Example 6.

Consider $U \sim \text{Ber}(p)$, $p \leq \frac{1}{2}$ and Hamming distortion. That is

$$d(u,v) = \begin{cases} 0 & \text{for } u = v \\ 1 & \text{for } u \neq v \end{cases}$$

Claim:

$$R(D) = \begin{cases} h_2(p) - h_2(D) & 0 \leq p \leq D \\ 0 & D > p \end{cases}$$

Proof: We will not be overly pedantic by worrying about small ϵ factors in the proof.

Note we can achieve distortion p without sending any information by setting $V = 0$. Therefore, for $D > p$, $R(D) = 0$, as claimed. For the remainder of the proof, therefore, we assume $D \leq p \leq \frac{1}{2}$.

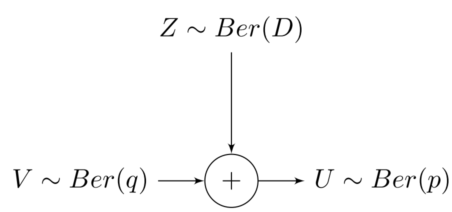
Consider U, V such that $U \sim \text{Ber}(p)$ and $\mathbb{E}d(U,V) = P(U \neq V) \leq D$. We can show that $R(D)$ is lower-bounded by $h_2(p) - h_2(D)$ by noting

$$\begin{aligned} I(U;V) &= H(U) - H(U|V) \\ &= H(U) - H(U \ominus_2 V|V) \\ &\geq H(U) - H(U \ominus_2 V) \\ &= h_2(p) - h_2(P(U \neq V)) \\ &\geq h_2(p) - h_2(D) \end{aligned}$$

In the second line we have used the fact that $H(U|V) = H(U \ominus_2 V|V)$ because there is a one to one mapping $(U,V) \leftrightarrow (U \ominus_2 V, V)$. In the third line, we have used that conditioning reduces entropy, so $H(U \ominus_2 V|V) \leq H(U \ominus_2 V)$. Finally, in the last line we have used that h_2 is increasing on $[0, \frac{1}{2}]$ and that $P(U \neq V) \leq D \leq p \leq \frac{1}{2}$. This establishes that $R(D) \geq h_2(p) - h_2(D)$.

Now we must show equality can be achieved. The first and second inequalities above demonstrate that we get equality if and only if

1. $U \ominus_2 V$ is independent of V.
2. $U \ominus_2 V \sim \text{Ber}(D)$.



Denoting $U \ominus_2 V \triangleq Z$, this is equivalent to finding q such that if $V \sim \text{Ber}(q)$ and $Z \sim \text{Ber}(D)$ is independent of V, $U = V \oplus_2 Z \sim \text{Ber}(p)$. Because $V \oplus_2 Z$ is binary, it is Bernoulli, with

$$\begin{aligned} p &= P(U = 1) \\ &= P(V = 1)P(Z = 0) + P(V = 0)P(Z = 1) \\ &= q(1 - D) + (1 - q)D \end{aligned}$$

Solving for q gives

$$q = \frac{p - D}{1 - 2D}$$

Because $D \leq p \leq \frac{1}{2}$, both the numerator and denominator are positive. Further, because $p \leq \frac{1}{2}$, we have $q \leq \frac{1/2-D}{1-2D} = \frac{1}{2}$, which shows that q is a valid probability. This completes the proof.

Example 7. Consider $U \sim \mathrm{N}(0, \sigma^2)$ and distortion given by: $d(u, v) = (U - V)^2$

Claim:

$$R(D) = \begin{cases} \frac{1}{2}\log((\sigma^2)/D) & 0 \leq D \leq p \\ 0 & D > \sigma^2 \end{cases}$$

Proof: First note we may achieve distortion σ^2 without transmitting any information by setting $V = 0$ with certainty. Therefore, $R(D) = 0$ for $D > \sigma^2$. For the remainder of the proof, therefore, we assume that $D \leq \sigma^2$.

For any U,V such that $U \sim \mathrm{N}(0, \sigma^2)$ and $\mathbb{E}(U - V)^2 \leq D$, we assume $D \leq \sigma^2$.

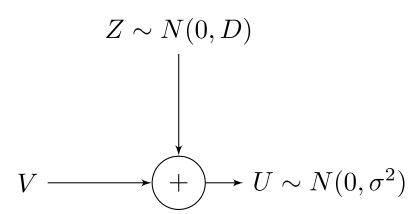
We can find the lower-bound by noting

$$\begin{aligned} I(U;V) &= h(U) - h(U|V) \\ &= h(U) - h(U - V|V) \\ &\geq h(U) - h(U - V) \\ &\geq h(U) - h(N(0, D)) \\ &= \frac{1}{2}\log 2\pi E\sigma^2 - \frac{1}{2}\log 2\pi e D \\ &= \frac{1}{2}\log\frac{\sigma^2}{D} \end{aligned}$$

For the first inequality we have used that conditioning reduces even the differential entropy, and in the second inequality we have used the result, proved earlier in the course that the maximum differential entropy of a distribution constrained by $\mathrm{Var}(U - V) \leq D$ is achieved when $U - V \sim \mathrm{N}(0, D)$. This establishes that $R(D) \geq \frac{1}{2}\log\frac{\sigma^2}{D}$.

Now we must show that equality can be achieved. The first and second inequalities above demonstrate that we get equality if and only if

1. $U - V$ is independent of V.
2. $U - V \sim \mathrm{N}(0, D)$.



Denoting $U - V \triangleq Z$, we want to find a distribution for V such that Z independent of Z and distributed $\mathrm{N}(0, D)$ makes $V + Z \sim \mathrm{N}(0, \sigma^2)$. We see that this is possible for $V \sim \mathrm{N}(0, \sigma^2 - D)$, which is a valid distribution because $D \leq \sigma^2$. This completes the proof, and $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$.