Information Theory with Applications, Math6397
Lecture Notes from October 21, 2014

taken by Junyu Ding

4 Rate distortion theory (continued)

We want to show for $\varepsilon > 0$, $R(D+\varepsilon) \leq \rho(D) + 4\varepsilon$. We had choosen $0 < \tau < 4\varepsilon$, m_n for sufficiently large n,

$$\rho(D) + \frac{\tau}{2} \leq \frac{1}{n} \ln m_n \leq \rho(D) + \tau$$

,

$$\delta = \min\left\{\frac{\tau}{8}, \frac{\varepsilon}{1 + 2\|d\|_\infty}\right\}$$

Step 1: Select m_n codewords $C_n = \{c_1, c_2, \ldots, c_{m_n}\}$ from $\hat{\mathbb{A}}^n$ randomly s.t

$$\forall \hat{a} \in C_n, \mathbb{P}_{\hat{X}^{\otimes n}}(\hat{a}) = \prod_{j=1}^{n} \mathbb{P}_{\hat{X}}(\hat{a}_j)$$

where each $\hat{X}_j$ is optimal, meaning $\rho(D) = I(X_j, \hat{X}_j)$ for each j.

Step 2: Given the codebook C_n and any $a \in \mathbb{A}^n$, choose

$$\phi_n(a) = \begin{cases} c, & \text{if there is } c \in C_n \text{ with } (a,c) \in \mathcal{D}_\delta^n \\ c_1, & else \end{cases}$$

Note: If $(a, \phi(a)) \in \mathcal{D}_\delta^n$, then by definition of $\mathcal{D}_\delta^n$,

$$\frac{1}{n} d_n(a, \phi_n(a)) \leq \mathbb{E}[d(X_1, \hat{X}_1)] + \delta$$

Next, we show that for our choice of ϕ, $\mathbb{E}[d(X_1, \hat{X}_1)] \leq D + \epsilon$.

Let us compute the probability of no match for a in distortion-typical set(mismatch).

We know $\mathbb{P}_{(X,\hat{X})^{\otimes n}}((\mathcal{D}_\delta^n)^c) < \delta$. Let $\rho_n = \left\{a \in \mathbb{A}^n : (a,c) \in \mathcal{D}_\delta^n,\ for\ some\ c\ in\ C_n\right\}$.

Probability of mismatch is $\mathbb{P}_m = \mathbb{P}_{X^{\otimes n}}((\rho_n)^c)$.

Expected value of $\mathbb{P}_n$ w.r.t. random coding is, using independence of our choice for each $c \in C_n$, with $1_{\mathcal{D}_\delta^n}$ the indicator function for $\mathcal{D}_\delta^n$

$$\mathbb{E}[\mathbb{P}_m] = \sum_{a \in \mathbb{A}^n} \mathbb{P}_{x^{\otimes n}}(a)(1 - \sum_{c \in \hat{\mathbb{A}}^n} \mathbb{P}_{\hat{x}}(c) 1_{\mathcal{D}_\delta^n}(a,c))^{m_n}$$

We estimated $\mathbb{P}_{\hat{x}^{\otimes n}}(c) \geq \mathbb{W}(c \mid a) e^{-n(I(X_1,\hat{X}_1)+3\delta)}$, so we get

$$(*)\mathbb{E}[P_m] \leq \sum_{a \in \mathbb{A}^n} \mathbb{P}_{X^{\otimes n}}(a)(1 - \sum_{c \in \hat{\mathbb{A}}^n} e^{-n(I(X_1,\hat{X}_1)+3\delta)} \mathbb{W}(c \mid a) 1_{\mathcal{D}_\delta^n}(a,c))^{m_n}$$

Interlude: If $0 \leq x \leq 1, 0 \leq y \leq 1, n \geq 1$, than $(1-xy)^m \leq 1 - x + e^{-ym}$

Proof: Fix y, then $(1-xy)^m$ is convex in x,

$$\begin{aligned}
(1-xy)^m &= (1 - ((1-x)(0) + x(1))y)^m \\
&\leq (1-x)(1)^2 + x(1-y)^m \\
&\leq 1 - x + xe^{-my} \\
&\leq 1 - x + e^{-my} .
\end{aligned}$$

Inserting this estimate in $(*)$, we get

$$\begin{aligned}
\mathbb{E}[P_m] &\leq \sum_{a \in \mathbb{A}^n} \mathbb{P}_{x^{\otimes n}}(1 - \sum_{c \in \hat{\mathbb{A}}^n} \mathbb{W}(c \mid a) 1_{\mathcal{D}_\delta^n}(a,c) + e^{-m_n e^{-n(I(X_1,\hat{X}_1)+3\delta)}}) \\
&(recall : \frac{1}{n} \ln m_n > \rho(D) + \frac{\tau}{2}) \\
&\leq \sum_{a \in \mathbb{A}^n} \mathbb{P}_{x^{\otimes n}}(1 - \sum_{c \in \hat{\mathbb{A}}^n} \mathbb{W}(c \mid a) 1_{\mathcal{D}_\delta^n}(a,c) + e^{-e^{n\rho(D)+\frac{\tau}{2}} e^{-n(I(x_1,\hat{x_1})+3\delta)}}) \\
&(recall : \delta = \min\left\{\frac{\tau}{8}, ...\right\}, so\ 4\delta \leq \tau/2)
\end{aligned}$$

so since ρ and I cancel

$$\begin{aligned}
\mathbb{E}[P_m] &\leq \sum_{a \in \mathbb{A}^n} \mathbb{P}_{x^{\otimes n}}(a)(1 - \sum_{c \in \hat{\mathbb{A}}^n} \mathbb{W}(c \mid a) 1_{\mathcal{D}_\delta^n}(a,c) + e^{-n\delta}) \\
&= 1 - \mathbb{P}_{(X,\hat{X})^{\otimes n}}(\mathcal{D}_\delta^n) + e^{-n\delta} \\
&\leq \delta + \delta = 2\delta
\end{aligned}$$

By the bound for $\mathbb{E}[P_m]$ (expectation w.r.t random choice of c), there exists C_n for which $\mathbb{P}_m \leq 2\delta$. For the average distortion, we then get

$$\begin{aligned}
\frac{1}{n}\mathbb{E}[d_n(x, \phi_n(x))] &= \sum_{a \in \mathbb{J}^A} \mathbb{P}_{x^{\otimes n}}(a) \frac{1}{n} d_n(a, \phi_n(a)) + \sum_{a \in (\mathbb{J}^A)^c} \mathbb{P}_{x^{\otimes n}}(a) \frac{1}{n} d_n(a, \phi_n(a)) \\
&\leq \sum_{a \in \rho_n} \mathbb{P}_{x^{\otimes n}}(a)(D + \delta) + \mathbb{P}_m \|d\|_\infty \\
&\leq D + \delta + 2\delta \|d\|_\infty (recall : \delta = \min\left\{..., \frac{\epsilon}{1 + 2\|d\|_\infty}\right\}) \\
&\leq D + \epsilon
\end{aligned}$$

We have just shown the existence of a code with exp. distortion $D + \epsilon$, and rate $\frac{1}{n} \ln m_n \to \rho(D) + 4\epsilon$, so $R(D + \epsilon) \leq \rho(D) + 4\epsilon$.

Step 3: Show the lower bound

We need to show that for any sequence (n, m_n, D) with $\limsup_n \frac{1}{n} \ln m_n < \rho(D)$, there is $\epsilon > 0$ and for sufficiently large n, $D_n = \frac{1}{n}\mathbb{E}[d_n(x, \phi_n(x))] > D + \epsilon$.

We observe

$$\begin{aligned}
\ln m_n &\geq H(\phi_n(X))(recall : \phi_n : deterministic) \\
&= H(\phi_n(X)) - H(\phi_n(X) \mid X) \\
&= I(X; \phi_n(X)) \\
&= H(X) - H(X \mid \phi_n(X)) \\
&= \sum_{j=1}^{n} H(X_j) - \sum_{j=1}^{n} H(X_j \mid \hat{X}_j, X_1, X_2, ..., X_{j-1}) \\
&= \sum_{j=1}^{n} H(X_j) - \sum_{j=1}^{n} H(X_j \mid \hat{X}_j) \\
&= \sum_{j=1}^{n} I(X_j, \hat{X}_j) \\
&\geq n \sum_{j=1}^{n} \frac{1}{n} \rho(D_j), with D_j = \mathbb{E}[d(X_j, \hat{X}_j)] \\
&\geq n\rho(\sum_{j=1}^{n} \frac{1}{n} D_j) \\
&= n\rho(\frac{1}{n}\mathbb{E}[d_n(X, \phi_n(X))]) = n\rho(\mathbb{E}[d(X_1, \hat{X}_1]) \,.
\end{aligned}$$