EE376A/STATS376A Information Theory Lecture 17 - 03/06/2018

Lecture 17: Strongly Typical Sequences and Rate Distortion 2

Lecturer: Tsachy Weissman *Scribe: yoann, genthial, cbournho, olivierm, rlenain*

1 Recap

A quick recap of our current setting.

$$U_1 \dots U_n, \text{ iid} \sim U \rightarrow \boxed{\text{Encoder}} \xrightarrow{J \in \{1 \dots M\}} \boxed{\text{Decoder}} \rightarrow V_1 \dots V_n$$
$$\text{Rate is } \frac{\log M}{n} \frac{\text{bits}}{\text{symbol}}$$
$$\text{Distortion is } \mathrm{d}(U^n, V^n) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{d}(U_i, V_i)$$

We say a pair (R, D) is achievable if for any $\epsilon > 0$, there exists a scheme (encoder, decoder pair) with rate less than or equal to $R + \epsilon$, and expected distortion $\mathrm{E}[d(U^n, V^n)] \leq D + \epsilon$. In this setting, we define

$$R(D) = \inf \{R' : (R', D) \text{ is achievable}\}$$

$R(D)$ can be thought of as the minimum number of bits per symbol needed to achieve expected distortion D. From this, we have presented our main theorem for this section before, which states

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

In order to prove this main theorem, we split it into the two following parts, which together are equivalent to the theorem.

$$\text{Direct part: } R(D) \leq R^{(I)}(D) \text{ (Proven already)}$$
$$\text{Converse part: } R(D) \geq R^{(I)}(D) \text{ (Remains to prove)}$$

We will recap the idea for the proof of the direct part (sometimes called achievability). First, we generate a codebook $\{V^n(1), \dots, V^n(M)\}$ iid $\sim V$. Then, for a given $1 \leq i \leq m$:

$$\mathrm{P}((U^n, V^n(i)) \text{ is jointly typical}) \approx 2^{-nI(U;V)} \tag{1}$$

We have proved that a direct implication from (1) is:

$$\mathrm{P}((U^n, V^n(i)) \text{ is jointly typical for some } 1 \leq i \leq M) \approx 1, \text{ provided } M = 2^{nR}, \text{ for } R > I(U; V)$$

It is then clear that in order to cover the typical set $T_\delta(U)$, we need a number of points greater than or equal to $\frac{2^{nH(U)}}{2^{nH(U|V)}} = 2^{nI(U;V)}$.

2 Proof of the converse part

Fix a scheme satisfying

$$\mathbb{E}\left[d(U^n, V^n)\right] \leq D$$

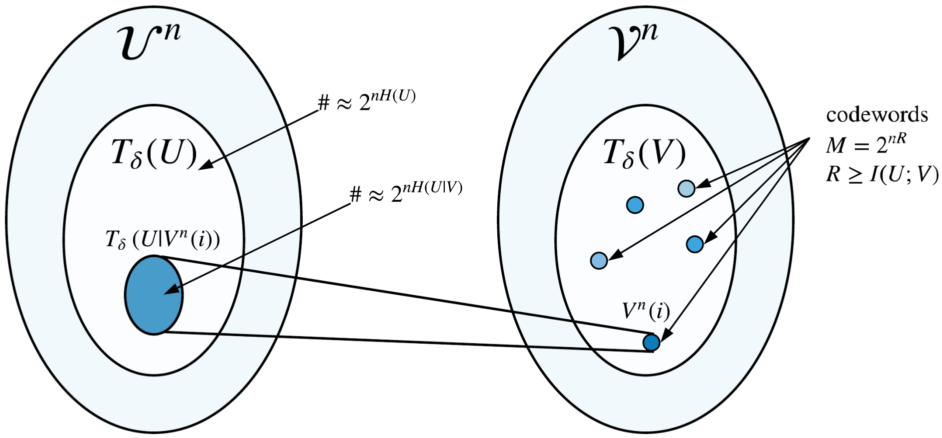

Figure 1: Typical sets and codewords - intuition. To cover $T_\delta(U)$ with conditionally typical balls, we need roughly $T_\delta(U)/T_\delta(U|V^n) \approx 2^{nI(U;V)}$ reconstruction sequences.

Then the entropy of the reconstruction under the scheme is no more than the log-size of the codebook, i.e.

$$H(V^n) \leq \log M$$

since the reconstruction takes values in a set of size at most M. Hence

$$\begin{aligned}
\log M &\geq H(V^n) \\
&\geq H(V^n) - H(V^n|U^n) \\
&= I(U^n; V^n) \\
&= H(U^n) - H(U^n|V^n) \\
&= \sum_{i=1}^{n} H(U_i) - H(U_i|U^{i-1}, V^n) \\
&\geq \sum_{i=1}^{n} H(U_i) - H(U_i|V_i) \quad \text{because conditioning reduces entropy} \\
&= \sum_{i=1}^{n} I(U_i; V_i) \\
&\geq \sum_{i=1}^{n} R^{(I)}\left(\mathbb{E}\left[d(U_i, V_i)\right]\right) \quad \text{from the definition of } R^{(I)}(D) \\
&\geq nR^{(I)}\Bigg(\underbrace{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[d(U_i, V_i)\right]}_{\mathbb{E}[d(U^n, V^n)] \leq D}\Bigg) \quad \text{from the convexity of } R^{(I)}(D) \text{ (homework 5)} \\
&\geq nR^{(I)}(D) \quad \text{from the monotonicity of } R^{(I)}(\cdot)
\end{aligned}$$

thus

$$\text{rate} = \frac{\log M}{n} \geq R^{(I)}(D)$$

which finishes the proof of the converse part

$$R(D) \geq R^{(I)}(D)$$

□

3 Example: Gaussian Source

In the case where $U \sim N(0, \sigma^2)$, $d(u, v) = (u - v)^2$.
Claim:

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

Which is equivalent to:

$$D(R) = \sigma^2 2^{-2R}$$

In particular, the best distortion that we can achieve if we are willing to dedicate only 1 bit per source symbol is $D(1) = \frac{\sigma^2}{4}$. This can be compared to the result at the start of class: $D = \frac{\pi-2}{\pi}\sigma^2 \approx 0.363\sigma^2$ (distortion that we can achieve if we work symbol by symbol and represent every symbol with one bit (see lecture note 2)).

Proof of claim:
For any U, V such that $U \sim N(0, \sigma^2)$ and $\mathbb{E}[(U - V)^2] \leq D$:

$$\begin{aligned} I(U;V) &= h(U) - h(U|V) \\ &= h(U) - h(U - V|V) \quad \text{differential entropy is invariant under a constant,and V is a constant given V} \\ &\geq h(U) - h(U - V) \quad \text{because conditioning reduces entropy} \\ &\geq h(U) - h(N(0, D)) \quad \text{gaussians maximize differential entropy among distributions} \\ &\qquad\qquad\qquad\qquad\quad \text{with bounded second moment} \\ &= \frac{1}{2}\log(2\pi e\sigma^2) - \frac{1}{2}\log(2\pi e D) \\ &= \frac{1}{2}\log\frac{\sigma^2}{D} \end{aligned}$$

$\Rightarrow R(D) \geq \frac{1}{2}\log\frac{\sigma^2}{D}$

We can achieve an equality if and only if:

1. $h(U - V|V) = h(U - V)$, i.e. $U - V$ independent from V
2. $h(U - V) = h(N(0, D))$, i.e. $U - V \sim N(0, D)$

Can we find a distribution V that satisfies these two conditions?
The answer is yes. If we take $V \sim N(0, \sigma^2 - D)$ and add an independent Gaussian $G \sim N(0, D)$, we reconstruct U by $U = V + G$.

1. $U - V = G$ is independent of V
2. $U - V = G \sim N(0, D)$

Conclusion: $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$ □

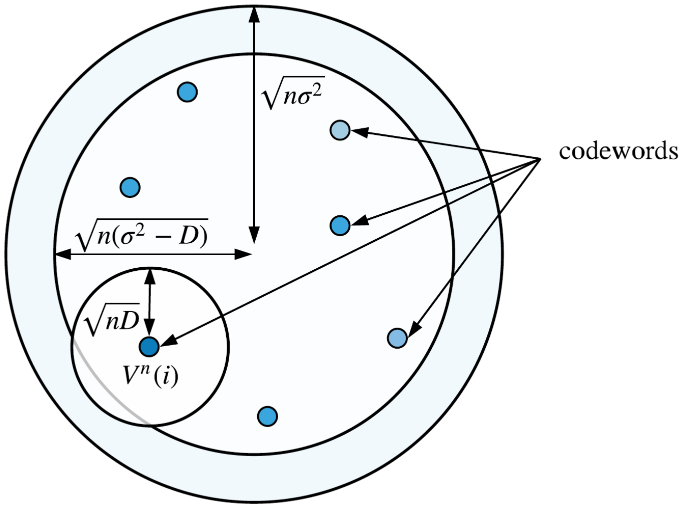

Figure 2: Gaussian example interpretation

4 Interpretation of the Gaussian example

We have found that the minimum achievable rate is $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$ when $U \sim N(0, \sigma^2)$.

If we want to visualize this, we can consider the values $U_1, ..., U_n$ as a vector in $\mathbb{R}^n$. Using the law of large numbers and the fact that $\mathbb{E}[U^2] = \sigma^2$, we know that:

$$\frac{1}{n}\sum_{i=1}^{n} U_i^2 \approx \sigma^2$$
$$\sum_{i=1}^{n} U_i^2 \approx n\sigma^2$$
$$\sqrt{\sum_{i=1}^{n} U_i^2} \approx \sqrt{n\sigma^2}$$
$$||U||_2 \lessapprox \sqrt{n\sigma^2}$$

So $(U_i) \in \mathbb{R}^n$ is in a ball centered on 0 and of radius $\sqrt{n\sigma^2}$.
Furthermore, we can represent each reconstructed element $V^n(i)$ for $i \in [1, M]$ of the codebook also in the space $\mathbb{R}^n$. As we need to achieve a distortion lower than D, each point $V^n(i)$ can represent points in a ball of radius $\sqrt{nD}$ centered on $V^n(i)$. This is because we have:

$$
\begin{aligned}
d(U^n, V^n) &\leq D \\
\frac{1}{n}\sum_{i=1}^{n}(U_i - V_i)^2 &\leq D \\
\frac{1}{n}||U - V||_2^2 &\leq D \\
||U - V||_2 &\leq \sqrt{nD}
\end{aligned}
$$

Therefore if we want to cover the whole ball of radius $\sqrt{n\sigma^2}$ with these small balls of radius $\sqrt{nD}$, we need the number of points in the codebook M to be:

$$
\begin{aligned}
M &\geq \frac{Vol(\text{ball of radius } \sqrt{n\sigma^2})}{Vol(\text{ball of radius } \sqrt{nD})} \\
&= \frac{c_n(\sqrt{n\sigma^2})^n}{c_n(\sqrt{nD})^n} \\
&= \left(\frac{\sigma^2}{D}\right)^{n/2}
\end{aligned}
$$

Because the rate is $R = \frac{m}{n} = \frac{\log M}{n}$, we obtain:

$$
\begin{aligned}
R &= \frac{\log M}{n} \\
&\geq \frac{1}{2}\log\frac{\sigma^2}{D}
\end{aligned}
$$

To rephrase, we need at least these M smaller balls to cover the full ball of radius $\sqrt{n\sigma^2}$. In lower dimensions, it looks like there is a lot of overlap between the smaller balls. But in higher dimensions, it is easy to cover the whole space in a very efficient way and achieve the optimal rate $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$.

The optimal V we found before is $V \sim N(0, \sigma^2 - D)$, which means we will take the reconstructed codewords $V^n(i)$ iid. on the sphere of radius $\sqrt{n(\sigma^2 - D)}$ to cover the whole space.

Even in low dimensions like $n = 5$ or $n = 6$ we can see that choosing these random codewords is already a very effective scheme. We will play with these schemes in the next homework.