EE376A/STATS376A Information Theory Lecture 2 - 01/11/2018

Lecture 2: Binary Sources, Lossy Compression and Channel Capacity

Lecturer: Tsachy Weissman *Scribe: Linda Banh, Josh Bosworth, Anthony Perez*

1 Example 2: Reliable Communication: Binary Source & Channel

1. **Source:** $\mathbb{U} = \{U_1, U_2, ...\}$ where $Pr[U_i = 0] = Pr[U_i = 1] = \frac{1}{2}$. The U_i's are i.i.d.
2. **Channel:** The channel flips each bit given to it with probability $q < \frac{1}{2}$. We define the channel input to be $\mathbb{X} = \{X_i\}$, the channel noise to be $\mathbb{W} = \{W_i\}$ and the channel output to be $\mathbb{Y} = \{Y_i\}$ such that:

$$W_i \sim Ber(q)$$
$$Y_i = X_i \oplus_2 W_i$$

   The W_i are i.i.d. and the X_i are functions of the input source sequence $\mathbb{U}$.
3. **Probability of error per source bit:** P_e, the probability of recovering U_i from Y_i.
4. **Rate:** the ratio $\frac{|\mathbb{U}|}{|\mathbb{X}|}$ or bits per channel use.

Encoding Scheme 1: The trivial encoding of $X_i = U_i$ yields a probability of error per source bit of $P_e = q$ because we decode each Y_i by assuming its value matches U_i. The rate for this scheme is 1 bit/channel use.

Encoding Scheme 2: We can repeat each source bit three times:

$$\mathbb{U} = 0\ 1\ 1\ 0\ ...$$
$$\mathbb{X} = 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ ...$$

This yields a probability of error per source bit of $P_e = 3q^2(1-q) + q^3 < q$ where the optimal decoding is to decode each Y_i by taking a majority vote amongst its three bits. However, this yields a rate of $\frac{1}{3}$ because the channel input is three times the size of the source.

Encoding Scheme 3: We can repeat each source bit K times (for simplicity, assume K is odd). The Rate becomes $\frac{1}{K}$ and the optimal decoding scheme remains to decode according to the majority bit. The probability of error is the probability that there are more bits which are corrupted than not-corrupted. q is our error rate, K is the total number of bits, and $\frac{K+1}{2}$ is the smallest number of corrupted bits that will result in an error.

$$P_e = \sum_{i=\frac{K+1}{2}}^{K} \binom{K}{i} (q)^i (1-q)^{K-i}$$

Thus, by increasing K, we can obtain a sequence of schemes for which the probability of error goes to 0 (can be shown using the law of large numbers since the number of errors converge to $Kq < K/2$). However, the rate for this sequence of schemes converges to 0 as $K \to \infty$. For a long time, it was believed that $P_e \to 0$ was possible only if rate $\to 0$. However,

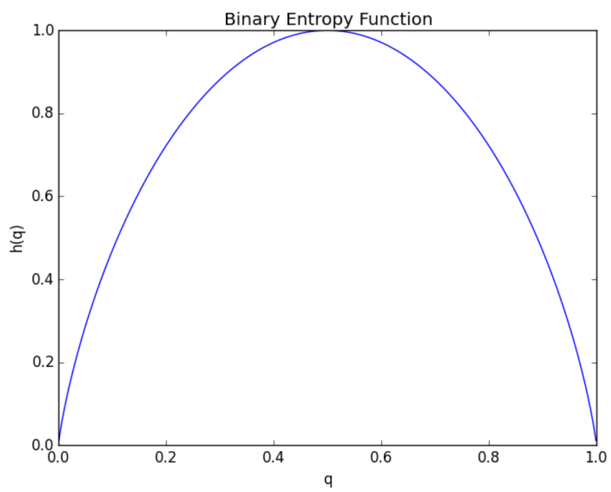
1.1 Theorem 1

Shannon 1948: $\exists\, R > 0$ and schemes with rate $\geq R$ satisfying $P_e \to 0$.

Definition 1. *Channel Capacity:*

$$C \triangleq \textit{Largest } R \textit{ for which Theorem 1 holds.}$$
$$C(q) = 1 - h(q)$$
$$h(q) \triangleq H(Ber(q)) = q \log \frac{1}{q} + (1-q) \log \frac{1}{1-q}$$

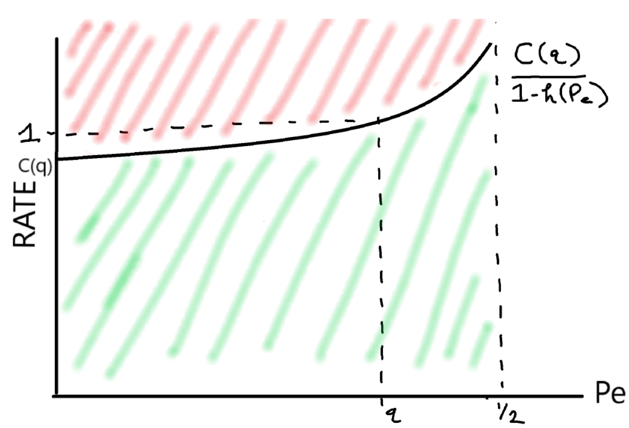
The figure below plots $h(q)$ for $q \in [0,1]$.



Note: Throughout the course, log will mean $\log_2$, unless otherwise specified. Also, $0 \log 0 \triangleq 0$.

1.2 Theorem 2

In general, if we are interested in a probability of error $P_e > 0$:
For any rate $< \frac{C(q)}{1-h(P_e)}$ where P_e is the probability of error we are willing to tolerate, there exists a scheme that achieves this probability of error and rate.
For any rate $> \frac{C(q)}{1-h(P_e)}$ no scheme can achieve the probability of error P_e.



The figure above shows the achievable and non-achievable pairs of rate and P_e in green and red, respectively. At $P_e = 0$, any rate below $C(q)$ is achievable. At $P_e = 0.5$, the achievable rate becomes infinite. This is because we can achieve probability of error 0.5 without transmitting any bits at all and randomly generating the bits at the receiver. Also, at $P_e = q$, the maximum rate is 1. This is because we can achieve $P_e = q$ by sending the source as it is without any coding (i.e., $X_i = U_i$). Thus, the simple scheme we did earlier is the optimal scheme if we are interested in a probability of error of q.

2 Example 3: Lossy Compression

General Objective: We are given a sequence of continuous random variables. Clearly it is impossible to represent them exactly with any finite number of bits. In general we want to represent them with as few bits as possible which allow us to reconstruct the original sequence with a low RMSE between the reconstruction and source sequence. Here, we consider the problem of representing a source with 1 bit/source symbol while minimizing the reconstruction RMSE.

1. **Source sequence:** $\mathbb{U} = \{U_1, U_2, ...\}$ where the U_i are i.i.d $\sim N(0, \sigma^2)$.
2. **Rate:** We have a desired compression rate K whose units are bits per source symbol.
3. **Encoding:** $\mathbb{B} = \{B_i\}$ is the sequence of bits used to encode the source.
4. **Reconstruction:** $\mathbb{V} = \{V_i\}$ is the reconstruction of the source sequence from the encoding. Minimizing the RMSE yields $V_i = E[U_i|B_i]$ as the optimal decoding.
5. **Distortion:** $\mathbb{D} = \{D_i\}$ where $D_i = E[(U_i - V_i)^2]$

Example 1: Suppose $K = 1$ and $B_i = \begin{cases} 1 \text{ if } U_i \geq 0 \\ 0 \text{ if } U_i < 0 \end{cases}$. Then we see that the optimal reconstruction and corresponding distortions are:

Reconstruction:

$$\begin{aligned} V_i = E[U_i|B_i = 1] &= \int_0^\infty x \, \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \, dx \\ &= \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^\infty \sqrt{u} \, e^{-\frac{u}{2\sigma^2}} \, \frac{1}{2\sqrt{u}} du \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-\frac{u}{2\sigma^2}} du \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} (-2\sigma^2 e^{-\frac{u}{2\sigma^2}})|_0^\infty \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} 2\sigma^2 = \sigma\sqrt{\frac{2}{\pi}} \end{aligned}$$

For $V_i = E[U_i|B_i = 0]$ the above proof holds with the limits reversed, yielding $V_i = \begin{cases} \sigma\sqrt{\frac{2}{\pi}} \text{ if } B_i = 1 \\ -\sigma\sqrt{\frac{2}{\pi}} \text{ if } B_i = 0 \end{cases}$.

Distortion:

$$D_i = E[(U_i - V_i)^2]$$
$$= \frac{1}{2}Var(U_i|B_i = 1) + \frac{1}{2}Var(U_i|B_i = 0)$$

(U_i is symmetrical for $B_i = 0$ and B_i =1)

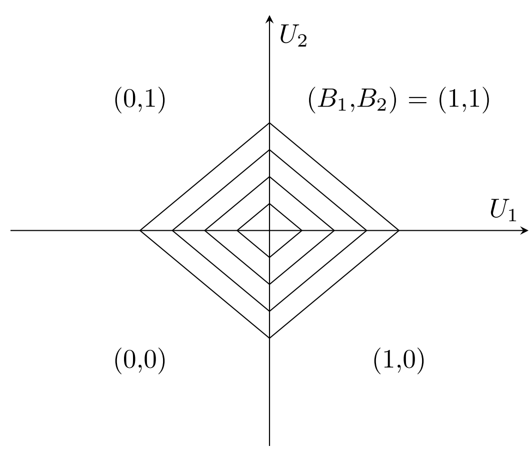
$$= Var(U_i|B_i = 1)$$
$$= E[U_i^2|B_i = 1] - (E[U_i|B_i = 1])^2$$

Using $E[U_i|B_i]$ that was calculated in **reconstruction:**

$$= \sigma^2 - (\sigma\sqrt{\frac{2}{\pi}})^2$$
$$= \sigma^2(1 - \frac{2}{\pi})$$
$$\approx 0.363\sigma^2$$

Example 2: As with the lossless compression example from Lecture 1, we can think about representing pairs of source symbols with pairs of bits.

Approach 1: Partitioning the plane into four quadrants

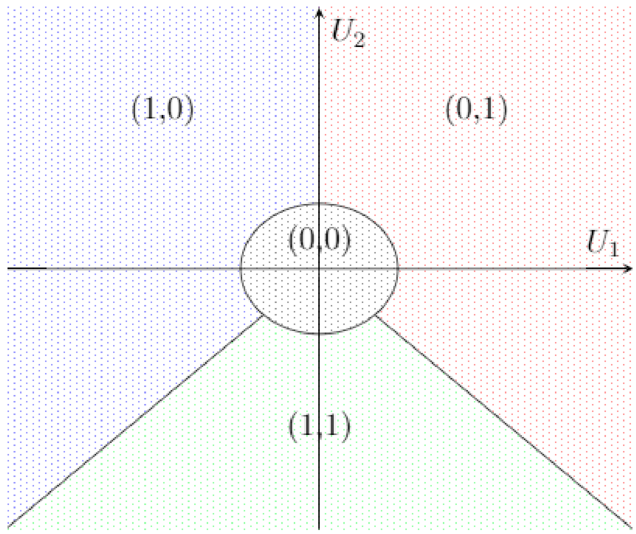


Observe that:

- U_1 is positive iff $B_1 = 1$
- U_2 is positive iff $B_2 = 1$
- U_1 is negative iff $B_1 = 0$
- U_2 is negative iff $B_2 = 0$

Thus, this scheme is exactly the same scheme as the symbol-by-symbol scheme discussed above. Hence, this will not offer any improvements in the distortion.

Approach 2: Partitioning the plane into a circle centered at the origin and into 3 symmetric regions



Exercise:

$$V_1(B_1, B_2) = E[U_1|B_1B_2] = ?$$
$$V_2(B_1, B_2) = E[U_2|B_1B_2] = ?$$

$$D_\rho = \frac{1}{2}E[(U_1 - V_1)^2 + (U_2 - V_2)^2]$$
$$\min_\rho D_\rho = ?$$

where ρ is the radius of the circular region. Although this is an interesting exercise, it cannot be easily solved in closed form. For the circular region, the reconstruction point is the origin (by symmetry). However, for the other regions, the integrals cannot be computed in closed form. Unfortunately, this happens quite commonly in higher dimensions. Therefore, it will be difficult to solve for D_ρ and to find the minimum ρ that minimizes that expression as well.

We will show later that: For any $\epsilon > 0$, $\exists$ schemes with $D \leq \sigma^2/4 + \epsilon$ using 1 bit/symbol. Also, for any scheme using 1 bit/symbol, $D \geq \sigma^2/4$. This gives a fundamental limit on lossy compression. Comparing this limit of $\sigma^2/4 = 0.25\sigma^2$ to the distortion for the first scheme ($0.363\sigma^2$), we see that there is some scope for improvement.

3 Example 4: Additive White Gaussian Noise Channel

For real world communication systems (e.g., wireless), the Additive White Gaussian Noise (AWGN) Channel is a natural model due to the central limit theorem.

1. **Information Source:** $\mathbb{U} = \{U_1, U_2, ...\}$. The U_i's are i.i.d. $\sim$ Ber($\frac{1}{2}$)
2. **Transmitted Signal:** $\mathbb{X} = \{X_1, X_2, ...\}$
3. **Received Signal:** $\mathbb{Y} = \{Y_1, Y_2, ...\}$

$$Y_i = X_i + N_i$$
$$N_i \sim N(0, \sigma^2) \ i.i.d$$

N_i's are also independent of $\mathbb{X}$.

4. **Rate:** Number of source bits transmitted per channel use:

$$Rate = N/n$$

5. **Channel Constraints:** We seek reliable communication, defined by a vanishing rate of error.

$$\lim_{n\to\infty} P(\hat{\mathbb{U}} \neq \mathbb{U}) = 0$$

We can trivially achieve this if we allow arbitrary large transmitted signals, so that the noise becomes insignificant in comparison. However, real world devices have power (P) constraints, and these power constraints limit the achievable rate.

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P$$

We'll show that we can achieve

$$Rate < \frac{1}{2}log(1 + P/\sigma^2)$$

and we cannot achieve

$$Rate > \frac{1}{2}log(1 + P/\sigma^2)$$

The Signal to Noise ratio is recognizable in the above equation as:

$$SNR = P/\sigma^2$$

In terms of the SNR, the expression for the capacity of the AWGN channel is $\frac{1}{2}log(1 + SNR)$.