1 Recap

- Source coding via asymptotic equipartition property (AEP)

- Jointly typicality of \((X^n, Y^n)\): The set \(A^{(n)}_\epsilon\) of jointly typical sequences \(\{(x^n, y^n)\}\)

\[
A^{(n)}_\epsilon = \{(x^n, y^n) \in X^n \times Y^n : \\
- \frac{1}{n} \log p(x^n) - H(X) < \epsilon, \\
- \frac{1}{n} \log p(y^n) - H(Y) < \epsilon, \\
- \frac{1}{n} \log p(x^n, y^n) - H(X,Y) < \epsilon \}
\]

- Joint AEP
  
  1. \(\Pr\left((X^n, Y^n) \in A^{(n)}_\epsilon\right) \to 1\) as \(n \to \infty\).
  2. \(\left|A^{(n)}_\epsilon\right| \leq 2^{n(H(X,Y)+\epsilon)}\).
  3. If \((\bar{X}^n, \bar{Y}^n) \sim p(x^n) p(y^n)\), then

\[
\Pr\left((\bar{X}^n, \bar{Y}^n) \in A^{(n)}_\epsilon\right) \leq 2^{-n(I(X;Y)-3\epsilon)}.
\]

- Channel capacity: \(C \triangleq \max_{p(x)} I(X;Y)\)

- Channel capacity of symmetric channel: \(C = \log |Y| - H(\text{row of transition matrix})\)

2 Channel Coding Theorem

- The most fundamental theorem in information theory

- Arguably, the first application of the probabilistic method in math

Definition 1 An \((M,n)\) code for the channel \((X, p(y \mid x), Y)\) consists of the following:
1. Message $W$ which is drawn from an index set $\{1, 2, \ldots, M\}$ (uniformly at random).

2. Encoder $X^n : \{1, \ldots, M\} \to X^n$. supp $(X^n)$ is a codebook. Each possible output is a codeword.

3. Decoder: deterministic function
   \[ g : Y^n \to \{1, \ldots, M\}. \]

**Definition 2 (Rate)** The rate $R$ of an $(M, n)$ code is
\[ R = \frac{\log M}{n} \text{ (bits per channel use)}. \]

**Definition 3 (Error probability)** Fix $i \in \{1, \ldots, M\}$,
\[ \lambda_i = \Pr (g(Y^n) \neq i \mid W = i), \]
\[ P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i = \Pr (W \neq g(Y^n)) \]
where $\Pr(W = i) = \frac{1}{M}$.

**Definition 4 (Achievable rate)** $R$ is achievable if there exists a sequence of codes at rate $\geq R$ such that $P_e^{(n)} \to 0$ as $n \to \infty$.

**Definition 5** $R^* \triangleq \sup \{R \mid R \text{ is achievable}\}$.

**Theorem 6 (Channel coding theorem)**
\[ R^* = C \]
which means that $C = \max_{p(x)} I(X; Y)$ is equal to the supremum of all achievable rates. All rates below capacity $C$ are achievable. Conversely, any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ must have $R \leq C$.

### 3 Achievability Proof

We prove that rates $R < C$ are achievable. If we fix some $R < C$ and $p(x)$, then there exist a code at rate $\geq R$ and its $P_e^{(n)}$ is arbitrarily small.

We create a random codebook and show it performs well. The random codebook $C$ is given by
\[ C = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(M) & x_2(M) & \cdots & x_n(M) \end{bmatrix} \]
where $M = 2^{nR}$ and we assume that $M$ is an integer.

We will consider the following.
• Each entry in $C$ is generated i.i.d. according to $p(x)$. Thus, the probability that we generate a particular codebook $C$ is $\Pr(C) = \prod_{w=1}^{2^nR} \prod_{i=1}^{n} p(x_i(w))$.

• Sender and receiver use this code (i.e., the codebook $C$ is revealed to both sender and receiver).

• Suppose a uniform random variable $W \in \{1, \ldots, M\}$ is sent.

• Receiver uses jointly typical decoding. Given $Y^n$, the decoder finds $X^n$ in the codebook such that $(X^n, Y^n)$ is jointly typical. If $X^n$ is unique, decode to $W$ corresponding to $X^n$. Else, declare an error (output can be a dummy value such as 0).

Let $\mathcal{E}$ be the error event.

\[
\Pr(\mathcal{E}) = \sum_{C} \Pr(C) P_{e}^{(n)}(C)
\]

\[
= \sum_{C} \Pr(C) \frac{1}{2^{nR}} \sum_{w=1}^{2^nR} \lambda_w(C)
\]

\[
= \frac{1}{2^{nR}} \sum_{w=1}^{2^nR} \sum_{C} \Pr(C) \lambda_w(C).
\]

Define the event $E_i$ that the $i$-th codeword is jointly typical with $Y^n$. Suppose that $W$ is fixed to $W = 1$. Then,

\[
\Pr(\mathcal{E} \mid W = 1) \leq \Pr(E_1^c \mid W = 1) + \sum_{i=2}^{2^nR} \Pr(E_i \mid W = 1)
\]

by the union bound. $E_1^c$ means that the transmitted codeword $X^n(1)$ and the received sequence $Y^n$ are not jointly typical.

Take $n$ large enough so that $\Pr(E_1 \mid W = 1) \geq 1 - \epsilon$ by joint AEP. Hence,

\[
\Pr(E_1^c \mid W = 1) \leq \epsilon.
\]

Since by the code generation process, $X^n(1)$ and $X^n(i)$ for $i \neq 1$ are independent, so are $Y^n$ and $X^n(i)$. By joint AEP,

\[
\Pr(E_i \mid W = 1) \leq 2^{-n(I(X;Y) - 3\epsilon)}.
\]

Consequently,

\[
\Pr(\mathcal{E} \mid W = 1) \leq \epsilon + \sum_{i=2}^{2^nR} 2^{-n(I(X;Y) - 3\epsilon)}
\]

\[
\leq \epsilon + 2^{3n\epsilon} 2^{-n(I(X;Y) - R)}
\]

\[
\leq 2\epsilon
\]

if $R < I(X;Y) - 3\epsilon$ and $n$ is large enough.

• This is true for all fixings of $W$. Hence, $\Pr(\mathcal{E}) \leq 2\epsilon$. 

3
• Choose $p(x) = p^*(x)$ that maximizes $I(X;Y)$. Then the condition $R < I(X;Y)$ can be replaced by the achievability condition $R < C$.

• Fix the code. It can be found by an exhaustive search.

• Get worst case. By Markov’s bound, $\Pr(\lambda_w \geq 4\epsilon) \leq \frac{P^{(n)}}{4\epsilon} = \frac{1}{2}$. Thus, at least half the indices $i$ and their associated codewords $X^n(i)$ must have conditional probability of error $\lambda_i$ less than $4\epsilon$. Hence the best half of the codewords have a maximal probability of error less than $4\epsilon$. Throwing away the worst half of the codewords (i.e., bad codewords) has changed the rate from $R$ to $R' = R - \frac{1}{n}$. If $n$ is large enough, $R' \approx R$. 