How to Connect Source Coding and Channel Coding

Suppose that $U$-valued source symbols arrive at a rate of one per unit time. If we collect source symbols over $n$ time units, we have $|U|^n$ possible source words. However, for a DMS with entropy $H(U)$, for large $n$, we only need to worry about $\exp(n[H(U) + \Delta R])$ of the source words.

Now consider a channel that sends one channel symbol per unit time. Suppose that $R > H(U)$ is an achievable channel-code rate; i.e., for large $n$, there are $\exp(n[R - \Delta R]) X^n$-valued codewords that can be transmitted with negligible error probability.

The key condition we require is that the number of compressed source messages be less than or equal to the number of codewords. Mathematically, we need

$$\left\lceil \exp(n[H(U) + \Delta R]) \right\rceil \leq \left\lfloor \exp(n[R - \Delta R]) \right\rfloor.$$

This will be true if

$$\exp(n[H(U) + \Delta R]) + 1 \leq \exp(n[R - \Delta R]) - 1,$$

and this holds for large $n$ as long as $R - \Delta R > H(U) + \Delta R$; i.e., if $0 < \Delta R < [R - H(U)]/2$. Notice that the condition $R > H(U)$ allows $\Delta R$ to be positive.

7.1.2. The Joint Asymptotic Equipartition Property

**Theorem 7.1** (Joint Asymptotic Equipartition Property (Joint AEP)). Let $A_n$ be defined as in Section 7.1 where $0 < \varepsilon < 1$ and $P_{XY}(x, y)$ have been given. The set of jointly typical pairs $A_n$ has small cardinality in the sense that

$$|A_n| \leq \exp(n[H(P_{XY}) + \varepsilon]).$$

For large $n$, the set $A_n$ has high $P_{XY}^n$ probability in the sense that

$$\lim_{n \to \infty} P_{XY}^n(A_n) = 1.$$

It then follows that for large $n$,

$$|A_n| \geq (1 - \varepsilon) \exp(n[H(P_{XY}) - \varepsilon]).$$
If \( Q_n(x, y) := P^n_X(x)P^n_Y(y), \) then

\[
Q_n(A_n) \leq \exp(-n[I(P_{XY}) - 3\varepsilon]),
\]

and for sufficiently large \( n, \)

\[
Q_n(A_n) \geq (1 - \varepsilon)\exp(-n[I(P_{XY}) + 3\varepsilon]).
\]

**Proof.** Because we have a lower bound on \( P_{XY}(x, y) \) for \((x, y) \in A_n\), we can write

\[
1 = \sum_{(x, y)} P^n_{XY}(x, y) \geq \sum_{(x, y) \in A_n} P^n_{XY}(x, y) \geq \sum_{(x, y) \in A_n} \exp(-n[H(P_{XY}) + \varepsilon])
\]

\[
= |A_n|\exp(-n[H(P_{XY}) + \varepsilon]).
\]

Rearranging yields the upper bound on \(|A_n|\).

To prove the limit result, it suffices to prove that \( 1 - P^n_{XY}(A_n) = P^n_{XY}(A^c_n) \to 0. \) Let \((X_1, Y_1), (X_2, Y_2), \ldots \) be i.i.d. pairs with common probability mass function \( P_{XY}. \) Put \( X := (X_1, \ldots, X_n) \) and \( Y := (Y_1, \ldots, Y_n). \) Then

\[
P((X, Y) \notin A_n) = P\left( \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{P_X(X_k)} - H(P_X) > \varepsilon, \right.
\]

or \( \left. \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{P_Y(Y_k)} - H(P_Y) > \varepsilon, \right) \)

or \( \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{P_{XY}(X_k, Y_k)} - H(P_{XY}) > \varepsilon \)

\[
\leq P\left( \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{P_X(X_k)} - H(P_X) > \varepsilon, \right.
\]

\[
+ P\left( \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{P_Y(Y_k)} - H(P_Y) > \varepsilon \right)
\]

\[
+ P\left( \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{P_{XY}(X_k, Y_k)} - H(P_{XY}) > \varepsilon \right).
\]

In each probability on the right, the terms in the sums are i.i.d., and the expectation of each logarithm is the corresponding entropy. Hence, by the weak law of large numbers, each probability tends to zero as \( n \to \infty. \)

Since \( P^n_{XY}(A_n) \to 1, \) for large \( n, \) we have \( P^n_{XY}(A_n) \geq 1 - \varepsilon. \) Hence,

\[
1 - \varepsilon \leq P^n_{XY}(A_n) = \sum_{(x, y) \in A_n} P^n_{XY}(x, y).
\]
\[
\leq \sum_{(x, y) \in A_n} \exp(-n[H(P_{XY}) - \varepsilon])
= |A_n| \exp(-n[H(P_{XY}) - \varepsilon]).
\]

Rearranging yields the lower bound on \(|A_n|\).

To establish the last two inequalities, write

\[
Q_n(A_n) = \sum_{(x, y) \in A_n} P^n_X(x) P^n_Y(y)
\leq \sum_{(x, y) \in A_n} \exp(-n[H(P_X) - \varepsilon]) \exp(-n[H(P_Y) - \varepsilon])
= |A_n| \exp(-n[H(P_X) - \varepsilon]) \exp(-n[H(P_Y) - \varepsilon])
\leq \exp(n[H(P_{XY}) + \varepsilon]) \exp(-n[H(P_X) - \varepsilon]) \exp(-n[H(P_Y) - \varepsilon])
= \exp(-n[I(P_{XY}) - 3\varepsilon]),
\]

and

\[
Q_n(A_n) = \sum_{(x, y) \in A_n} P^n_X(x) P^n_Y(y)
\geq \sum_{(x, y) \in A_n} \exp(-n[H(P_X) + \varepsilon]) \exp(-n[H(P_Y) + \varepsilon])
= |A_n| \exp(-n[H(P_X) + \varepsilon]) \exp(-n[H(P_Y) + \varepsilon])
\geq (1 - \varepsilon) \exp(n[H(P_{XY}) - \varepsilon]) \exp(-n[H(P_X) + \varepsilon]) \exp(-n[H(P_Y) + \varepsilon])
= (1 - \varepsilon) \exp(-n[I(P_{XY}) + 3\varepsilon]).
\]

\[\square\]

### 7.2. The Typicality Decoder for the DMC

Recall that in ML decoding, the sets \(B_i\) were defined only in terms of the channel transition probability \(W_n\). In **typicality decoding**, the sets \(B_i\) depend not only on \(W_n\) but also on a probability mass function \(p\) on \(X\). In the next section, we use \(p\) to construct a random codebook, but for now, \(p\) is just a parameter used to construct the \(B_i\).

For a DMC, recall that \(W_n(y|x) = W^n(y|x) := \prod_{k=1}^n W(y_k|x_k)\). The typicality decoder is defined in terms of the set,\(^b\)

\[
B_i := \{y: (x_i, y) \in A_n\},
\]

where now we take \(P_{XY}(x, y) := p(x)W(y|x)\) in the definition of \(A_n\) (Section 7.1). In other words, the decoder decides message \(i\) was sent if \(x_i\) and \(y\) are jointly typical. If

\(^b\) When we use a random codebook, \(B_i = \{y: (X_i, y) \in A_n\}\) depends on the random codeword \(X_i\).
if \( y \) is jointly typical with more than one codeword, we decide in favor of the smallest value of \( i \). If there is no codeword with which \( y \) is jointly typical, then we announce message \( N \).

### 7.3. Analysis of Error Probability

In our analysis, we take the random codewords \( X_i \) to be i.i.d. We further take the individual letters of each codeword to be i.i.d. Specifically, 

\[
P(X_i = x) = p^n(x).
\]

Next, we recall (5.8) and write

\[
W_n(\{y : \varphi(y) \neq i\} | x_i) \leq W_n(B^c_i | x_i) + W_n\left( \bigcup_{j \neq i} B_j | x_i \right).
\]

Using this bound on the right in (6.8), we get

\[
E[W_n(\{y : \varphi(y) \neq i\} | X_i)] \leq T_1 + T_2,
\]

where

\[
T_1 := \sum_x E\left[ W_n(B^c_i | X_i) \right] p^n(x),
\]

and

\[
T_2 := \sum_x E\left[ W_n\left( \bigcup_{j \neq i} B_j | X_i \right) \right] p^n(x).
\]

We now show that \( T_1 \) tends to zero as \( n \to \infty \). Write

\[
T_1 = \sum_x E\left[ \sum_{y : (x,y) \notin A_n} W_n(y | X_i) \right] p^n(x)
\]

\[
= \sum_x E\left[ \sum_{y : (x,y) \notin A_n} W_n(y | x) \right] p^n(x)
\]

\[
= \sum_x \sum_{y : (x,y) \notin A_n} W_n(y | x) p^n(x).
\]

Since we are working with a DMC, \( W_n(y | x) = W^n(y | x) \). Hence, \( W_n(y | x) p^n(x) = W^n(y | x) p^n(x) = P^n_{XY}(x,y) \). In other words, \( T_1 = P^n_{XY}(A_n^c) \), which tends to zero by the joint AEP.