

**Lecture 4 Supplemental Notes:**

In this lecture, we discussed the Asymptotic Equipartition Property (A.E.P.) for an i.i.d. sequence of discrete random variables. The A.E.P. tells us that

$$-\frac{1}{n} \log p(X_1, \dots, X_n) \rightarrow H(X) \quad \text{in probability,}$$

or equivalently,

$$\lim_{n \rightarrow \infty} \Pr((X_1, \dots, X_n) \in A_\epsilon^{(n)}) = 1,$$

where  $A_\epsilon^{(n)}$  denotes the *typical set* of sequences of length  $n$ . Following the text, we then showed several properties of the typical set.

These properties have a direct interpretation in terms of source coding. Recall, from Lec. 1, that information sources are modeled probabilistically. The simplest such class of models are called Discrete Memoryless Sources (DMS). The output of a DMS is simply an i.i.d sequence of RV's. The sample space for each random variable is referred to as the *source alphabet*.

Assume the source alphabet consists of  $|\mathcal{X}|$  letters. To label all length  $n$  sequences of source symbols with a unique, fixed-length binary string requires  $\lceil n \log_2 |\mathcal{X}| \rceil$  bits ( $\lceil x \rceil$  denotes the smallest integer  $\geq x$ ). Such a labeling is an example of a source code, this is defined more precisely in Chap. 5. The labels are referred to as codewords. This code is *fixed-length* since each label has the same number of bits and *lossless* since each source sequence is assigned a unique label.

Two possible ways to reduce the number of bits used on average are:

- Use a lossy source code; that is do not assign labels to all sequences of source letters.
- Use a variable-length source code, *i.e.* assign strings of different lengths to different sequences of source letters.

The A.E.P. tells us something about both of these. The second case is discussed in Theorem 3.2.1 in the text; This is an example of a source coding theorem. The following gives the analogous theorem for the case of fixed-length lossy source codes. Following the common convention, we divide this into two parts - a direct part and a converse:

**Source coding theorem - direct part:** Given a DMS with entropy  $H(X)$ , then for all  $\epsilon > 0$  and all  $p > 0$ , there exists an integer  $N > 0$  and a fixed length source code that represent sequences of  $N$  source symbols using  $NH(X) + \epsilon$  bits with probability of decoding error less than  $p$ .

**Converse:** For any  $\epsilon > 0$  and  $p \in (0, 0.5)$  then for large enough  $N$ , any source code that uses  $(H(X) - \epsilon)N$  bits to represent sequences of  $N$  source symbols will have a probability of decoding error that is larger than  $p$ .

The direct part or achievability is a direct consequence of the AEP. The converse follows from Theorem 3.3.1 in the book.

A stronger converse can be proven that shows that the probability of decoding error for a rate  $H(X) - \epsilon$  code approaches 1 for  $N$  large enough (see Section 3.1 of *Information Theory and Reliable Communication* by R. Gallager).