Iowa State University Lecturer: Aditya Ramamoorthy
EE 520: Topics in Communications: Multiuser Information Theory
August 26, 2008 Scribe: Aditya Ramamoorthy

Lecture 1

The concept of typical sequences is very fundamental to information theory. In this lecture we shall define different notions of typicality of sequences, that shall be used in later classes. We shall use the notation x^n to denote the sequence $x_1, x_2, \ldots, x_n$. For most of this class we shall be dealing with sources and channels that take values in a discrete alphabet.

1 Entropy-Typical seqences

Let x^n be a finite sequence, such that $x_i \in \mathcal{X}$, where $\mathcal{X}$ denotes a discrete alphabet. Let $N(a|x^n)$ be the number of positions of x^n having letter a, where $a \in \mathcal{X}$. Let $P_X^n(x^n) = \Pi_{i=1}^n P_X(x_i)$ denote the probability that the sequence x^n was emitted by a discrete memoryless source (DMS) $P_X(\cdot)$. Then,

$$P_X^n(x^n) = \begin{cases} \Pi_{a \in \mathrm{supp}(P_X)} P_X(a)^{N(a|x^n)} & \text{if } N(a|x^n) = 0 \text{ whenever } P_X(a) = 0. \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Intuitively, letter a occurs $\approx nP_X(a)$ times when n is large, so that

$$\begin{aligned} P_X^n(x^n) &\approx \Pi_{a \in \mathrm{supp}(P_X)} P_X(a)^{nP_X(a)}, \\ \implies -\frac{1}{n} \log_2 P_X^n(x^n) &\approx -\sum a \in \mathrm{supp}(P_X) P_X(a) \log_2 P_X(a) = H(X), \end{aligned}$$

where $H(X)$ denotes the entropy of X. Shannon's notion of typical sequences (sometimes referred to as weak-typicality) is the following

Definition 1. *Entropy-typical sequence. A sequence x^n is said to be typical with respect to an $\epsilon > 0$ and $P_X(\cdot)$ if*

$$\left| -\frac{1}{n} \log_2 P_X^n(x^n) - H(X) \right| < \epsilon.$$

Note that this is equivalent to,

$$2^{-n[H(X)+\epsilon]} < P_X^n(x^n) < 2^{-n[H(X)-\epsilon]}.$$

This notion of typicality is only concerned with the probability of the sequence and not the actual sequence itself. Next we define a stronger notion of typicality, called letter-typicality.

2 Letter-typical sequences

Here we say that a sequence is typical if its empirical pmf is close to the pmf from which it is generated. More formally, we have the following definition.

Definition 2. *A sequence x^n is said to be ϵ-letter typical with respect to $P_X(\cdot)$ if*

$$\left|\frac{1}{n}N(a|x^n) - P_X(a)\right| \leq \epsilon P_X(a),\ \forall a \in \mathcal{X} \tag{2}$$

The set of sequences that are ϵ-letter typical for a given n will be denoted by $T_\epsilon^n(P_X)$. For this class we shall mainly work with the letter-typicality notion. We now show that this notion is stronger than entropy-typicality.

Example. If $P_X(\cdot)$ is uniform over $\mathcal{X}$ then ϵ-letter typical x^n satisfy

$$\frac{(1-\epsilon)n}{|\mathcal{X}|} \leq N(a|x^n) \leq \frac{(1+\epsilon)n}{|\mathcal{X}|} \quad \text{when } \epsilon < |\mathcal{X}| - 1.$$

i.e. not all x^n are ϵ-letter typical. However, all x^n have $P_X^n(x^n) = \frac{1}{|\mathcal{X}|^n}$ i.e. all x^n are entropy-typical since $P_X^n(x^n) = 2^{-nH(X)}$. Thus, we have shown an example of a sequence that is entropy-typical but is not ϵ-letter typical. Next, we show that any sequence that is ϵ-letter typical is also entropy-typical.

Let $\mu_X = \min_{x\in \text{supp}(P_X)} P_X(x)$ and $\delta_\epsilon(n) = 2|\mathcal{X}|e^{-n\epsilon^2\mu_X}$ and note that $\delta_\epsilon(n) \to 0$ as $n \to 0$.

Theorem 1. *Suppose that $0 \leq \epsilon \leq \mu_X$, $x^n \in T_\epsilon^n(P_X)$ and X^n is emitted by a DMS, $P_X(\cdot)$. We have,*

i) $2^{-n(1+\epsilon)H(X)} \leq P_X^n(x^n) \leq 2^{-n(1-\epsilon)H(X)}$.

ii) $(1-\delta_\epsilon(n))2^{n(1-\epsilon)H(X)} \leq |T_\epsilon^n(P_X)| \leq 2^{n(1+\epsilon)H(X)}$.

iii) $1 - \delta_\epsilon(n) \leq P[X^n \in T_\epsilon^n(P_X)] \leq 1$.

For large n and small ϵ, the intuition for these results is as follows. The first results states that the probability of typical sequences is concentrated tightly around $2^{-nH(X)}$. The second result says that there are approximately $2^{nH(X)}$ sequences in the typical set $T_\epsilon^n(P_X)$ and the third result states that with high probability any sequence emitted by the DMS is typical.

Proof. Let $x^n \in T_\epsilon^n(P_X)$. To show (i), we proceed as follows.

$$\begin{aligned} P_X^n(x^n) &= \Pi_{a\in\text{supp}(P_X)} P_X(a)^{N(a|x^n)} \\ &\leq \Pi_{a\in\text{supp}(P_X)} P_X(a)^{n(1-\epsilon)P_X(a)} \text{ using letter-typicality definition, eqn. (2)} \\ &= 2^{n(1-\epsilon)\sum_{a\in\text{supp}(P_X)} P_X(a)\log_2 P_X(a)} \\ &= 2^{-n(1-\epsilon)H(X)}. \end{aligned}$$

The other direction can be proved similarly.

Next we prove (iii).

$$\begin{aligned} P[X^n \notin T_\epsilon^n(P_X)] &= P\left[\cup_{a\in\text{supp}(P_X)} \left\{\left|\frac{1}{n}N(a\middle|x^n) - P_X(a)\right| > \epsilon P_X(a)\right\}\right] \\ &\leq \sum_{a\in\text{supp}(P_X)} P\left(\left|\frac{1}{n}N(a|x^n) - P_X(a)\right| > \epsilon P_X(a)\right). \end{aligned}$$

Now using Chernoff bounds, it is possible to show that

$$P\left(\left|\frac{1}{n}N(a|x^n) - P_X(a)\right| > \epsilon P_X(a)\right) \leq 2e^{-n\epsilon^2 \mu_X},$$

where $0 \leq \epsilon \leq \mu_X$. Using this we have the required result.

To show that (ii) holds, we note that

$$\begin{aligned} \sum_{x^n \in T_\epsilon^n(P_X)} P_X^n(x^n) &\leq 1 \\ \implies |T_\epsilon^n(P_X)| 2^{-n(1+\epsilon)H(X)} &\leq 1 \quad \text{from part (i)} \\ \implies |T_\epsilon^n(P_X)| &\leq 2^{n(1+\epsilon)H(X)} \end{aligned}$$

The other direction can be proved in a similar fashion.