R. Gallager, Information Theory and Reliable Communication

EE725 Φυλλάδιο 1
9/3/2011

Since (3.5.9) is valid for all L, we have

$$H_\infty(U) \le \lim_{L\to\infty} \; (U_L | \, U_1 \cdots U_{L-1}) \tag{3.5.10}$$

Equations 3.5.10 and 3.5.3 together establish (3.5.4), completing the proof. |

Theorem 3.5.2. (Variable-Length Source Coding Theorem). Let $H_L(U)$ be the entropy per letter in a sequence of length L for a discrete source of alphabet size K. Given a code alphabet of D symbols, it is possible to encode sequences of L source letters into a prefix condition code in such a way that the average number of code letters per source letter, $\bar{n}$, satisfies

$$\frac{H_L(U)}{\log D} \le \bar{n} < \frac{H_L(U)}{\log D} + \frac{1}{L} \tag{3.5.11}$$

Furthermore, the left-hand inequality must be satisfied for any uniquely decodable set of code words for the sequences of L source letters. Finally, for any $\delta > 0$, if the source is stationary it is possible to choose L large enough so that $\bar{n}$ satisfies

$$\frac{H_\infty(U)}{\log D} \le \bar{n} < \frac{H_\infty(U)}{\log D} + \delta \tag{3.5.12}$$

and $\bar{n}$ can never violate the left-hand inequality for any uniquely decodable code.

Proof. The proof of (3.5.11) is the same as that of Theorem 3.3.2, except that $LH_L(U)$ is the entropy of a sequence of L source letters rather than $LH(U)$. Taking the limit in (3.5.11) as $L \to \infty$, $H_L(U)$ approaches $H_\infty(U)$ and $1/L$ approaches 0, establishing (3.5.12). |

In our discussion of discrete memoryless sources, our interest in $\bar{n}$ was motivated by the law of large numbers, which asserted that the number of code letters per source letter in a long sequence of code words approaches $\bar{n}$. The following example shows that this limiting behavior need not hold for arbitrary discrete stationary sources. Suppose that a source with the alphabet (a_1,a_2,a_3) has two modes of behavior, each occurring with probability ½. In the first mode the source produces an infinite sequence of repetitions of a_1. In the second mode, the source products an infinite sequence of statistically independent, equiprobable selections of the letters a_2 and a_3. If we encode sequences of L source letters into a binary code, it is not hard to see that $\bar{n}$ is minimized by mapping the sequence of a_1's into a single binary digit and mapping each of the 2^L sequences of a_2's and a_3's into code words of length $L + 1$. Since the mode of the source never changes, either all code words in a sequence will be of length 1 or all will be of length $L + 1$. For such sources, neither $\bar{n}$ nor the entropy are quantities of any great significance.

Sources that cannot be separated into different persisting modes of behavior are known as ergodic sources. To define ergodicity carefully, let $\mathbf{u} = \ldots, u_{-1}, u_0, u_1, \ldots$ be an infinite length sequence of source letters and let $T^l\mathbf{u}$ denote the sequence $\mathbf{u}$ shifted in time by l positions. That is, if we denote $T^l\mathbf{u}$ by $\mathbf{u}'$ we have

$$u_n' = u_{n+l}; \qquad -\infty < n < \infty$$

Likewise, if S is a set of infinite length sequences of source letters, T^lS denotes the same set shifted by l positions. That is, if $\mathbf{u}' = T^lu$, then $\mathbf{u}'$ is in the set T^lS iff $\mathbf{u}$ is in S. A set of sequences is said to be *invariant* if $TS = S$. It can easily be seen that the set of all sequences from a discrete source is invariant, and also that for any $\mathbf{u}$, the set $\cdots T^{-1}\mathbf{u}, \mathbf{u}, T\mathbf{u}, T^2\mathbf{u}, \ldots$ is invariant. *A discrete stationary source is defined to be ergodic if every measurable, invariant set of sequences has either probability one or probability zero.* It can be seen that, in the previous example, the set of sequences in each of the modes of behavior referred to are invariant sets, each with probability ½. Thus that source is nonergodic.

The above definition, although elegant, is sometimes difficult to work with and does not bring out the intuitive concept of ergodicity. An equivalent definition is as follows. Let $f_n(\mathbf{u})$ be a function of the infinite length source sequence $\mathbf{u}$ which depends only on a finite sequence, $u_1, \ldots, u_n$, of the source letters. Then a discrete stationary source is *ergodic* iff for all $n \ge 1$ and all $f_n(\mathbf{u})$ for which $|\overline{f_n(\mathbf{u})}| < \infty$, we have, for all source sequences $\mathbf{u}$ except a set of probability 0,

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} f_n(T^l\mathbf{u}) = \overline{f_n(\mathbf{u})} \tag{3.5.13}$$

The class of functions in the definition can be expanded to all measurable functions $f(\mathbf{u})$ for which $|\overline{f(\mathbf{u})}| < \infty$, or can be restricted to the special class of functions $f_{\mathbf{u}_n'}(\mathbf{u})$ where $\mathbf{u}_n'$ is a fixed sequence $u_1', \ldots, u_n'$ of letters and

$$f_{\mathbf{u}_n'}(\mathbf{u}) = \begin{cases} 1 & \text{if } u_1 = u_1',\, u_2 = u_2', \ldots, u_n = u_n' \\ 0 & \text{otherwise} \end{cases} \tag{3.5.14}$$

For a proof of the equivalence of these definitions, see Khinchin (1957), pp. 49–54, and for yet another equivalent definition, see Wolfowitz (1961), Lemma 10.3.1.

The definition of (3.5.13) is particularly important since it is the result that we shall need for ergodic sources. What it says is that the law of large numbers applies to ergodic sources. Alternatively, it says that the time average, averaged in time over any sample source output (except a set of zero probability), is equal to the ensemble average $\overline{f_n(\mathbf{u})}$. Since $\overline{f_{\mathbf{u}_n'}(\mathbf{u})}$ is just the probability of sequence $\mathbf{u}_n'$, (3.5.14) states that the relative frequency of occurrence of $\mathbf{u}_n'$ in a very long source sequence will be approximately equal to the probability of $\mathbf{u}_n'$.

Unfortunately, not even ergodicity quite implies that the number of code letters per source letter in a variable length code approaches $\bar{n}$. If we encode L source letters at a time and let $n(u_1, \ldots, u_L)$ be the length of a code word, then the time average number of code letters per source letter is given by

$$\lim_{J \to \infty} \frac{1}{JL} \sum_{j=0}^{J-1} n(u_{Lj+1}, \ldots, u_{Lj+L}) \tag{3.5.15}$$

See Problem 3.21 for an example of an ergodic source where this average, as a random variable, takes on different values with nonzero probability. The difficulty is that (3.5.15) is not a time average in the same sense as (3.5.13) since it is defined in terms of shifts of L letters at a time rather than shifts of one letter at a time.

Fortunately, Theorem 3.1.1 does apply to arbitrary ergodic sources. The major difficulty in proving this lies in establishing a law of large numbers for self-information; that is, in showing that $I(\mathbf{u}_L)/L$ is, with high probability, close to $H_\infty(U)$ for large L. This law of large numbers is of considerable mathematical and information theoretic interest and we now state it as a theorem.

Theorem 3.5.3 (McMillan (1953)). Let a discrete stationary ergodic source have $H_1(U) < \infty$.

For arbitrary $\epsilon > 0$, $\delta > 0$, there exists an integer $L_o(\epsilon,\delta)$ (which depends upon the source) such that for all $L \geq L_o(\epsilon,\delta)$

$$\Pr\left[\left|\frac{I(\mathbf{u}_L)}{L} - H_\infty(U)\right| > \delta\right] < \epsilon \tag{3.5.16}$$

Before proving the theorem, we shall develop some necessary notation and prove two lemmas. We observe that

$$\frac{I(\mathbf{u}_L)}{L} = \frac{1}{L} \sum_{l=1}^{L} I(u_l \mid u_1, \ldots, u_{l-1}) \tag{3.5.17}$$

Notice that the right-hand side of (3.5.17) closely resembles a time average as given in (3.5.13). The difference is that each self-information term in (3.5.17) depends on a different number of previous source digits. The point of the proof is to show that this dependence dies out sufficiently quickly as l becomes large. Let $P(\mathbf{u}_L) = P(u_1, \ldots, u_L)$ denote the probability assignment on a sequence of L source digits, and define $Q_m(\mathbf{u}_L)$, for any integer $1 \leq m \leq L$ by

$$Q_m(\mathbf{u}_L) = P(\mathbf{u}_m) \prod_{l=m+1}^{L} P(u_l \mid u_{l-1}, \ldots, u_{l-m}) \tag{3.5.18}$$