CHAPTER 13
Source Coding Revisited — Discrete Stationary Sources

13.1. Stationarity and the Entropy Rate

Let $X$ be a finite set. Let $\{X_k\}_{k=-\infty}^{\infty}$ be a family of $X$-valued RVs such that for $n = 1, 2, \ldots$, and all integers $k$,

$$P(X_{k+1} = x_1, \ldots, X_{k+n} = x_n) = P(X_1 = x_1, \ldots, X_n = x_n),$$

where $x_1, \ldots, x_n$ are in $X$. In this case, the process $\{X_k\}$ is said to be (strictly) stationary. For such process we show later that

(a) $H(X_n|X_1, \ldots, X_{n-1})$ is nonincreasing with $n$.
(b) $\frac{1}{n}H(X_1, \ldots, X_n) \geq H(X_n|X_1, \ldots, X_{n-1})$.
(c) $\frac{1}{n}H(X_1, \ldots, X_n)$ is nonincreasing with $n$.
(d) $\lim_{n\to\infty} \frac{1}{n}H(X_1, \ldots, X_n) = \lim_{n\to\infty} H(X_n|X_1, \ldots, X_{n-1})$.

The fact that entropy is bounded below by zero combined with properties (a) and (c) implies that these limits exist.

The entropy rate of a stationary process $\{X_k\}$ is

$$\mathcal{H} := \lim_{n\to\infty} \frac{1}{n}H(X_1, \ldots, X_n).$$

We then have the following generalization of the Variable-Length Source Coding Theorem 3.17 (Note: If $\{X_k\}$ are i.i.d. with common pmf $p$, then $\mathcal{H} = H(p)$.)

**Theorem 13.1.** Let $X$ be a finite set, and let $\{X_k\}_{k=-\infty}^{\infty}$ be a stationary $X$-valued source. Then for every blocklength $n \geq 1$, there is a prefix code $C_n:X^n \to D^*$ such that

$$\frac{1}{n}H(X_1 \cdots X_n) \log D \leq \frac{E[\ell(C_n(X_1, \ldots, X_n))]}{n} < \frac{1}{n}H(X_1 \cdots X_n) \log D + \frac{1}{n}.$$
Furthermore, given $\varepsilon > 0$, for all sufficiently large $n$,

$$\frac{\mathcal{H}}{\log D} \leq \frac{E[\ell(C_n(X_1, \ldots, X_n))]}{n} < \frac{\mathcal{H}}{\log D} + \varepsilon.$$ 

### 13.2. Ergodic Sources

A stationary process $\{X_k\}$ is said to be **ergodic** if it has the additional property that for every function $f(x)$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = E[f(X_1)], \quad \text{a.s.,}$$

whenever $E[f(X_1)]$ exists. An i.i.d. sequence is stationary, and by the **strong law of large numbers**, the sequence is ergodic.

**Example 13.2** (A Stationary Process That Is Not Ergodic). Let $Z$ be a binary random variable taking the values 0 and 1 with equal probability. Put $X_k := Z$ for all $k$. Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k = Z,$$

which is either zero or one, and is therefore never equal to $E[X] = E[Z] = 1/2!$

### 13.3. Typical Sequences Again

For $\Delta R > 0$, put

$$A_n := \left\{ (x_1, \ldots, x_n) \in X^n : \left| \frac{1}{n} \log \frac{1}{p_n(x_1, \ldots, x_n)} - \mathcal{H} \right| \leq \Delta R \right\},$$

where $p_n(x_1, \ldots, x_n) := P(X_1 = x_1, \ldots, X_n = x_n)$. Then $x = (x_1, \ldots, x_n) \in A_n$ implies

$$\exp(-n[\mathcal{H} + \Delta R]) \leq p_n(x) \leq \exp(-n[\mathcal{H} - \Delta R]).$$

With $X = (X_1, \ldots, X_n)$, it would be nice to have $\lim_{n \to \infty} P(X \notin A_n) = 0$. If the $X_k$ were i.i.d., we could use the weak law of large numbers. For stationary ergodic processes, we use the following result.
Theorem 13.3 (Shannon–McMillan–Breiman). Let $X$ be a finite set, and let $\{X_k\}$ be a stationary, ergodic, $X$-valued process. Then

$$\lim_{n \to \infty} P((X_1, \ldots, X_n) \notin A_n) = 0.$$  

Proof. See below. 

With the Shannon–McMillan–Breiman theorem, it is straightforward to adapt the proof of Shannon’s Source Coding Theorem 4.3 to prove the following result.

Theorem 13.4. Let $X$ be a finite set, and let $\{X_k\}_{k=-\infty}^{\infty}$ be a stationary, ergodic, $X$-valued source. Then every $R \geq H$ is achievable, and every $R < H$ is not achievable.


The remainder of this section is devoted to proving the Shannon–McMillan–Breiman theorem.

Lemma 13.5. Let $X$ be a finite set, and let $\{X_k\}$ be a stationary, ergodic, $X$-valued process. For $m \geq 1$,

$$\lim_{n \to \infty} \frac{1}{n} \log Q_m(X_1, \ldots, X_n) = -H(X_{m+1}|X_m, \ldots, X_1), \quad \text{a.s.},$$

where for $n > m$,

$$Q_m(x_1, \ldots, x_n) := P(X_1 = x_1, \ldots, X_m = x_m) \cdot \prod_{k=m+1}^{n} P(X_k = x_k|X_{k-1} = x_{k-1}, \ldots, X_{k-m} = x_{k-m}).$$

Proof. Write,

$$\frac{1}{n} \log Q_m(X_1, \ldots, X_n) = \frac{1}{n} \log p_m(x_1, \ldots, x_m)$$

$$+ \frac{1}{n} \sum_{k=m+1}^{n} \log p_{X_k|X_{k-1}, \ldots, X_{k-m}}(X_k|X_{k-1}, \ldots, X_{k-m}).$$

Clearly, $\frac{1}{n} \log p_m(x_1, \ldots, x_m) \to 0$ a.s. as $n \to \infty$. Since

$$E[\log p_{X_k|X_{k-1}, \ldots, X_{k-m}}(X_k|X_{k-1}, \ldots, X_{k-m})] = -H(X_k|X_{k-1}, \ldots, X_{k-m})$$

$$= -H(X_{m+1}|X_m, \ldots, X_1),$$

124
and since the source is ergodic, the term
\[
\frac{1}{n} \sum_{k=m+1}^{n} \log P_{X_k|X_{k-1},\ldots,X_{k-m}}(X_k|X_{k-1},\ldots,X_{k-m}) \to -H(X_{m+1}|X_m,\ldots,X_1)
\]
almost surely.