ENEE 627
SPRING 2011
INFORMATION THEORY

ENTROPY RATES OF SOURCES

Throughout $\mathcal{X}$ is a finite set. Consider an information sourcce modeled by the $\mathcal{X}$-valued sequence $\mathbb{X} = \{X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. We shall write

$$p_n(\boldsymbol{x}^n) = \mathbb{P}\left[X_1 = x_1, \ldots, X_n = x_n\right]$$

for each $\boldsymbol{x}^n = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$. Thus, $\boldsymbol{p}_n = (p_n(\boldsymbol{x}^n,\ \boldsymbol{x}^n \in \mathcal{X}^n)$ is the pmf of the random vector $\boldsymbol{X}^n = (X_1, \ldots, X_n)$.

Defining entropy rates

The entropy rate $H(\mathbb{X})$ of the sequence $\mathbb{X} = \{X_n,\ n = 1, 2, \ldots\}$ is defined by

$$H(\mathbb{X}) = \lim_{n \to \infty} \frac{H(X_1, \ldots, X_n)}{n} \tag{1}$$

provided this limit exists. Note that this definition is well adapted for the statement of the Source Coding Theorem.

As seen in class, this definition is well posed in many cases of interest, including i.i.d. sequences, Markov chains and stationary sequences. For a stationary sequence, it is possible to show that the convergence (1) is always guaranteed and that it takes place monotonically with

$$\frac{H(X_1, \ldots, X_n, X_{n+1})}{n+1} \leq \frac{H(X_1, \ldots, X_n)}{n}, \quad n = 1, 2, \ldots$$

This is an easy consequence of the chain rule for entropies and of properties of Cesaro convergence with

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | \boldsymbol{X}^{i-1}), \quad n = 1, 2, \ldots$$

Asymptotic Equipartition Property (AEP)

In many instances there exists a *non*-random constant $H^\star(\mathbb{X})$ such that the convergence

$$-\frac{1}{n}\log_2 p_n(X_1,\ldots,X_n) \xrightarrow{P}{}_n H^\star(\mathbb{X}) \tag{2}$$

takes place. This convergence in probability amounts to

$$\lim_{n\to\infty} \mathbb{P}\left[\left|-\frac{1}{n}\log_2 p_n(X_1,\ldots,X_n) - H^\star(\mathbb{X})\right| > \varepsilon\right] = 0. \tag{3}$$

In particular, this is equivalent to the following: For every $\varepsilon > 0$ and $\delta > 0$, there exists a finite integer $n^\star(\varepsilon,\delta)$ such that

$$\mathbb{P}\left[\left|-\frac{1}{n}\log_2 p_n(X_1,\ldots,X_n) - H^\star(\mathbb{X})\right| > \varepsilon\right] \leq \delta, \quad n \geq n^\star(\varepsilon,\delta) \tag{4}$$

This is often applied with $\delta = \varepsilon$ (in which case we write $n^\star(\varepsilon) = n^\star(\varepsilon,\varepsilon)$).

This convergence gives rise to the Asymptotic Equipartition Property (AEP) to be stated shortly: For $\varepsilon > 0$, set

$$A_n(\varepsilon) \equiv \left\{ \boldsymbol{x}^n \in \mathcal{X}^n : \left|-\frac{1}{n}\log_2 p_n(x_1,\ldots,x_n) - H^\star(\mathbb{X})\right| > \varepsilon \right\},$$

for each $n = 1,2,\ldots$. A sequence $\boldsymbol{x}^n$ in $A_n(\varepsilon)$ is said to be an ε-weak typical sequence of length n for the source $\mathbb{X}$.

Lemma 0.1 *Assume the convergence (2) to hold for some non-random constant $H^\star(\mathbb{X})$. Then, for every $\varepsilon > 0$, the following statements are true:*

(i) The convergence

$$\lim_{n\to\infty} \mathbb{P}\left[(X_1,\ldots,X_n) \in A_n(\varepsilon)\right] = 1 \tag{5}$$

holds;

(ii) The upper bound

$$|A_n(\varepsilon)| \leq 2^{n(H^\star(\mathbb{X})+\varepsilon)} \tag{6}$$

holds for all $n = 1,2,\ldots$; and

(ii) The lower bound

$$(1-\varepsilon)2^{n(H^\star(\mathbb{X})-\varepsilon)} \leq |A_n(\varepsilon)| \tag{7}$$

holds for all $n = 1, 2, \ldots$ *sufficiently large.*

The AEP (and its variants) are key for establishing the Channel Coding Theorem.

$H(\mathbb{X})$ **vs.** $H^\star(\mathbb{X})$

A natural question arising from these definitions is whether $H^\star(\mathbb{X})$ and $H(\mathbb{X})$ always coincide when both are well defined. While in general this is not so, the following relationship holds.

Lemma 0.2 *Assume* $H(\mathbb{X})$ *and* $H^\star(\mathbb{X})$ *to be well defined according to (1) and (2), respectively. It is always the case that*

$$H^\star(\mathbb{X}) \leq H(\mathbb{X}). \tag{8}$$

The existence of a possible relationship between the quantities $H^\star(\mathbb{X})$ and $H(\mathbb{X})$ is already suggested by the following observation:. For each $n = 1, 2, \ldots$, we have

$$\begin{aligned} & \mathbb{E}\left[-\log_2 p_n(X_1, \ldots, X_n)\right] \\ &= -\sum_{\boldsymbol{x}^n \in \mathcal{X}^n} \log_2 p_n(\boldsymbol{x}^n) \cdot p_n(\boldsymbol{x}^n) \\ &= H(X_1, \ldots, X_n) \end{aligned} \tag{9}$$

so that

$$\mathbb{E}\left[-\frac{1}{n}\log_2 p_n(X_1, \ldots, X_n)\right] = \frac{H(X_1, \ldots, X_n)}{n}. \tag{10}$$

In view of (1) and (2), it is tempting to let n go to infinity in this last equality. Were we able to exchange limits and expectations, we would establish the equality $H^\star(\mathbb{X}) = H(\mathbb{X})$. Unfortrunately additional conditions are needed in order to achieve this interchange.

Proof. Recall that convergence in probability implies a.s. convergence along some subsequence. Thus, in view of the convergence (2) there exists a subsequence $k \to n_k$ with $\lim_{k\to\infty} n_k = \infty$ such that

$$\lim_{k\to\infty} \left(-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right) = H^\star(\mathbb{X}) \quad a.s. \tag{11}$$

Note that under this subsequence, the definition of $H(\mathbb{X})$ also implies via (1) that

$$H(\mathbb{X}) = \lim_{k\to\infty} \frac{H(X_1, \ldots, X_{n_k})}{n_k} \tag{12}$$

With these preliminaries in mind, consider the equality (10) along the subsequence used in (11). Fatou's Lemma yields

$$\begin{aligned} & \mathbb{E}\left[\liminf_{k\to\infty} \left(-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right) \right] \\ \le \quad & \liminf_{k\to\infty} \mathbb{E}\left[-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right] \end{aligned} \tag{13}$$

by the non-negativity

$$-\frac{1}{n} \log_2 p_n(X_1, \ldots, X_n) \ge 0, \quad n = 1, 2, \ldots$$

Using (11) we conclude that

$$\begin{aligned} & \mathbb{E}\left[\liminf_{k\to\infty} \left(-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right) \right] \\ = \quad & \mathbb{E}\left[\lim_{k\to\infty} \left(-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right) \right] \\ = \quad & H^\star(\mathbb{X}). \end{aligned} \tag{14}$$

On the other hand, it follows from (10) and (12) that

$$\begin{aligned} & \liminf_{k\to\infty} \mathbb{E}\left[-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right] \\ = \quad & \lim_{k\to\infty} \mathbb{E}\left[-\frac{1}{n_k} \log_2 p_{n_k}(X_1, \ldots, X_{n_k}) \right] \\ = \quad & H(\mathbb{X}). \end{aligned} \tag{15}$$

Combining (14) and (15) with (13) gives the desired conclusion (8). ■