Information Theory with Applications, Math6397
Lecture Notes from September 18, 2014

taken by Nicole Leonhard

2.3.6 Remarks (Homework 1, Problem 3 Solution). Idea: Look at the sum of binomial coefficients and then look at what that means for counting. Then use the hint provided.

Proof. Let $0 < \alpha \leq \frac{1}{2}$ and $\{X_1, \ldots, X_n\}$ be binary random variables such that each outcome $\omega \in \Omega$, with $X_1 + \cdots X_n \leq \alpha n$, is equally probable. In this case, entropy is maximized and its value is

$$H(X_1, \cdots, X_n) = \ln|\Omega| = \ln\left(\sum_{i\in I}\binom{n}{i}\right)$$

Now, using additivity, we have

$$H(X_1, \ldots, X_n) \leq \sum_{j=1}^{n} H(X_j) = \sum_{j=1}^{n} H(X_1) = nH(X_1)$$

Each X_j is a binary random variable. Recall entropy for binary random variables of success probability p, $h(p)$ has a maximum at p=1/2. Next, we wish to estimate p. To this end, we introduce the sum variable $S = X_1 + X_2 + \cdots + X_n$. Conditioning on the value of $S = \lfloor \alpha n \rfloor$, we have $\mathbb{P}(X_1 = 1 | S = \lfloor \alpha n \rfloor) \leq \alpha$ and conditioning on the lower values for S gives strictly smaller success probabilities, so averaging over all the possible outcomes for S shows that $p = \mathbb{P}(X_1 = 1) < \alpha$. So, by the monotonicity of $h(x)$ for $p \leq \alpha \leq \frac{1}{2}$ we get $H(X_1) < h(\alpha)$ and thus

$$\ln\left(\sum_{i\in I}\binom{n}{i}\right) < nh(\alpha).$$

Exponentiate and we get the claimed inequality. □

Recall from last time we discussed:

- Block coding theorem and the converse
- Stationary ergodic processes and Birkhoff's Theorem.

Last time, in Birkhoff's Ergodic Theorem we had almost sure convergence. This makes the Strong law of Large numbers a special case of Birkhoff's Ergodic Theorem. We will now show the convergence in Birkhoff's Ergodic Theorem is stronger than the convergence in probability we were using in the first result on asymptotic equipartitioning. In brief, almost sure-convergence implies convergence in probability.

To show this, assume that for a sequence of random variables $\{S_n\}_{n=1}^{\infty}$ and random variable S we have a set

$$E = \left\{\omega \in \Omega : \lim_{n\to\infty} S_n(\omega) \text{ exists and is equal to S}\right\}$$

with measure $\mathbb{P}(E) = 1$. Now, consider $\epsilon > 0$. Let

$$U_n = \{\omega \in \Omega | S_n(\omega) - S(\omega) < \epsilon, \text{for all} N \geq n\}.$$

Since the $\lim_{n\to\infty} S_n(\omega)$ exists, we know that $E \subset \cup_{n\in N} U_n$ and thus $\mathbb{P}(\cup_{n\in N} U_n) = 1$. By the regularity of $\mathbb{P}$ and the fact that $U_1 \subset U_2 \subset \cdots$, we have $\lim_{n\to\infty} \mathbb{P}(U_n) = 1$ which implies

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : |S_n(\omega) - S(\omega)| < \epsilon\}) = 1$$

which is convergence in probability, as appearing in the weak law of large numbers.

Now we want use the ideas from discrete memoryless sources for block codes for stationary ergodic sources. But, now we cannot just look at X_1 anymore because the X_j's could be dependent.

2.3.7 Definition. Given a source $\{X_j\}_{j=1}^{\infty}$ with at most countable alphabet, its entropy rate is

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

provided the limit exists.

We will see that this limit exists for stationary sources.

2.3.8 Lemma. *For a stationary source, $\{X_j\}_{j=-\infty}^{\infty}$, the conditional entropy, $H(X_n|X_{n-1}, \ldots, X_1)$ has a limit.*

Proof. Let $\{X_j\}_{j=-\infty}^{\infty}$ be a stationary source. From the definition of conditional entropy, we know $H(X_n|X_{n-1}, \ldots, X_1) \geq 0$ and

$$\begin{aligned} H(X_n|X_{n-1}, \ldots, X_1) \leq & H(X_n|X_{n-1}, \ldots, X_2) && \text{(Since conditioning decreases entropy)} \\ & = H(X_{n-1}|X_{n-2}, \ldots, X_1) && \text{(Since the source is stationary)} \end{aligned}$$

So, a sequence of conditional entropies of a stationary source is non increasing. This together with the fact that it is bounded below gives us that it has a limit.

□

2.3.9 Lemma (Cesaro means)**.** *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers and* $b_n = \frac{1}{n}\sum_{j=1}^{n} a_j$*. If* $a_n \to a$ *as* $n \to \infty$*, then* $b_n \to a$ *as* $n \to \infty$*.*

Proof. Let $\epsilon > 0$. By the convergence of $\{a_n\}_{n=1}^{\infty}$ to a, there exists $N \in \mathbb{N}$ such that for all $n > N$, we have $|a_n - a| < \epsilon$. Thus,

$$|b_n - a| = |\frac{1}{n}\sum_{j=1}^{n} a_j - a| = |\frac{1}{n}\sum_{j=1}^{n} a_j - \frac{1}{n}\sum_{j=1}^{n} a| =$$

$$\frac{1}{n}\sum_{j=1}^{n} |a_j - a| = \frac{1}{n}\sum_{j=1}^{N} |a_j - a| + \frac{1}{n}\sum_{j=N+1}^{n} \underbrace{|a_j - a|}_{which\ is\ <\epsilon\ by\ assumption}$$

$$\leq \frac{1}{n}\sum_{j=1}^{N} |a_j - a| + \frac{1}{n}(n - N)\,\epsilon.$$

So, we have

$$0 \leq \limsup_{n\to\infty} |b_n - a| < \lim\sup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{N} |a_j - a| + \frac{1}{n}(n - N)\,\epsilon = 0 + \epsilon$$

This holds for all $\epsilon > 0$. Therefore, $\lim_{n\to\infty} |b_n - a| = 0$. □

2.3.10 Theorem. *For a discrete stationary source,* $\{X_j\}_{j=1}^{\infty}$*, the entropy rate exists and is given by*

$$\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1).$$

Proof. By lemma 2.3.8, $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$ exists. By additivity,

$$\frac{1}{n} H(X_1, \ldots, X_n) = \frac{1}{n}\sum_{j=1}^{n} H(X_j | X_{j-1}, \ldots, X_1)$$

. So $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$ is a sequence of Cesaro means for conditional entropy. By the convergence of the conditional entropies, the Cesaro means also converge to the same limit. □

2.3.11 Theorem (Shannon-McMillan-Breaiman)**.** *Let* $\{X_j\}_{j=-\infty}^{\infty}$ *be a stationary ergodic source with at most countable alphabet, then* $-\frac{1}{n}\ln\left(\mathbb{P}_{X_1,\ldots,X_n}(X_1, \ldots, X_n)\right)$ *converges to*

$$\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1),$$

with probability 1.

This statement is stronger than the one from last time, since almost sure convergence implies convergence in probability.

Sketch of proof. For simplicity of notation, we denote $\mathbb{P}_{X_1,\ldots,X_N}$ as $\mathbb{P}$. We have

$$\mathbb{P}\left(X_1,\ldots,X_n\right)=\mathbb{P}\left(X_1\right)\prod_{j=2}^{n}\mathbb{P}\left(X_j|X_{j-1},\ldots,X_1\right).$$

So,

$$\begin{aligned}\frac{1}{n}\ln\mathbb{P}\left(X_1,\ldots,X_n\right)&=\frac{1}{n}\ln\mathbb{P}\left(X_1\right)\left(\prod_{j=2}^{n}\mathbb{P}\left(X_j|X_{j-1},\ldots,X_1\right)\right)\\&=\frac{1}{n}\left(\ln\mathbb{P}\left(X_1\right)+\sum_{j=2}^{n}\ln\left(\mathbb{P}\left(X_j|X_{j-1},\ldots,X_1\right)\right)\right).\end{aligned}$$

Now, we can compare the series term by term with the expected values,

$$\frac{1}{n}H\left(X_1,\ldots,X_n\right)=\frac{1}{n}\sum_{j=1}^{n}H\left(X_j|X_{j-1},\ldots,X_1\right)$$

.

Using the Martingale convergence theorem it can be shown that

$$\lim_{n\to\infty}\ln\left(\mathbb{P}\left(X_j|X_{j-1},\ldots,X_1\right)\right)$$

exists almost surely. When considering

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\ln\left(\mathbb{P}\left(X_j|X_{j-1},\ldots,X_1\right)\right)$$

the average is approximately over shifted copies, which would allow us to use Birkhoff's theorem. Indeed, a detailed study shows that

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\ln\underbrace{\left(\mathbb{P}\left(X_j|X_{j-1},\ldots,X_1\right)\right)}_{\text{Notice these all have the same limit}}&=\\ \text{(Since our source is stationary ergodic)}&=\mathbb{E}\left[\ln\left(\mathbb{P}\left(X_j|X_{j-1},X_{j-2},\ldots\right)\right)\right]\\ \text{(Since our source is stationary we can replace the j with n.)}&=\lim_{n\to\infty}H\left(X_n|X_{n-1},\ldots\right)\\ \text{(From the previous theorem,)}&=\lim_{n\to\infty}\frac{1}{n}H\left(X_1,\ldots,X_n\right),\end{aligned}$$

Which proves the theorem.

□

2.3.12 Remarks. For alternative proofs, see R.M Gray's book "Entropy and Information Theory", which can be viewed at http://ee.stanford.edu/∼gray/it.pdf. The details begin on page 50.