

Information Theory with Applications, Math6397

Lecture Notes from September 16, 2014

taken by John Haas

Last Time

- Data processing for Markov chains
- Asymptotic Equipartition Principle For Shannon and McMillan
 - For large n , A_n^ϵ has overwhelming probability
 - Each outcome in A_n^ϵ is approximately equally likely with probability between $e^{-n(H(X_1)-\epsilon)}$ and $e^{-n(H(X_1)+\epsilon)}$
 - Block codes

Sources and Coding, Continued

2.3 Block Codes For DMS, Continued

In the last lecture, we presented the asymptotic equipartition principle (AEP), which roughly states that when given a DMS and sufficiently large n , then an outcome sequence $x = \{x_1, x_2, \dots, x_n\}$ is most likely occur within the "important set" A_n^ϵ and with approximately the same probability as any other $y \in A_n^\epsilon$. This result is the key to proving the following results about reconstruction under block coding. We concluded the last lecture by stating the *Block Coding Theorem*, but we did not have time to present proof, so we resume today's lecture with a proof of this theorem. For the readers' convenience, we restate the theorem again here.

2.3.7 Theorem. (*Block Coding Theorem*) Let $\{X_j\}_{j=1}^\infty$ be a DMS with entropy $H(X_1)$ and let $\epsilon > 0$, then there exists $\delta \in (0, \epsilon)$ and a sequence of codes $\{(\mathcal{C}_n, \phi_n)\}_{n=1}^\infty$ with block sizes $\{m_n\}_{n=1}^\infty$ (i.e., $|e_n| = m_n$), and there exist maps $\{\psi_n\}_{n=1}^\infty$, $\psi_n : \mathcal{C}_n \rightarrow \mathbb{A}^n$ such that

$$\frac{1}{n} \ln(m_n) < H(X_1) + \delta$$

and

$$\mathbb{P}(\psi_n \circ \phi_n(X_1, \dots, X_n) \neq (X_1, \dots, X_n)) < \epsilon$$

for all sufficiently large n .

Proof. Fix $\delta \in (0, \epsilon)$. Identify elements in $A_{\delta/2}^n$ with codewords in a set \mathcal{C}_n^* in such a way that there is a 1 – 1 map from $A_{\delta/2}^n$ to \mathcal{C}_n^* , then extend this 1 – 1 map to a map, ϕ_n , defined on all of \mathbb{A}^n by setting $\phi_n(x) = c_0$ for all $x \notin A_{\delta/2}^n$ and let \mathcal{C}_n be the disjoint union $\mathcal{C}_n = \mathcal{C}_n^* \dot{\cup} \{c_0\}$. Now choose n large enough so that $\frac{n\delta}{2} > \ln(2)$ and so that $\mathbb{P}((x_1, \dots, x_n) \notin A_{\delta/2}^n) < \frac{\delta}{2}$ (as in the Shannon-McMillan-Breiman theorem from the last lecture). This yields

$$\begin{aligned} m_n &= |A_{\delta/2}^n| + 1 \\ &< e^{n(H(X_1) + \delta/2)} + 1 && \text{(by A.E.P)} \\ &< 2e^{n(H(X_1) + \delta/2)} \\ &< e^{n(H(X_1) + \delta)} && \text{(because } 2 = e^{\ln(2)} < e^{n\frac{\delta}{2}} \text{)} \end{aligned}$$

and hence

$$\frac{1}{n} \ln(m_n) < H(X_1) + \delta.$$

Finally, consider the map which is the left inverse of the restriction $\phi_n|_{A_{\delta/2}^n}$ and extend this to the map, ψ_n , defined on all of \mathcal{C}_n by setting $\psi_n(c_0) = x$ for ANY choice of $x \in \mathbb{A}^n$. This gives us

$$\begin{aligned} \mathbb{P}(\psi_n \circ \phi_n(X_1, \dots, X_n) \neq (X_1, \dots, X_n)) &\leq \mathbb{P}((X_1, \dots, X_n) \notin A_{\delta/2}^n) \\ &< \frac{\delta}{2} \\ &< \epsilon \end{aligned}$$

as desired. □

Now we provide a converse to this theorem, which states that reconstruction fails with overwhelming probability when the codebooks are not sufficiently large. Once again, this result depends heavily on the asymptotic equipartitioning principle.

2.3.8 Theorem. (Converse To Block Coding) *If $\{(\mathcal{C}_n, \phi_n)\}_{n=1}^\infty$ is a sequence of codes with codebook sizes $\{m_n\}_{n=1}^\infty$ such that $\limsup_{n \rightarrow \infty} \frac{1}{n} \ln(m_n) < H(X_1)$, then for any $\lambda > 0$ and any sequence of pairs $\{(\phi_n : \mathbb{A}^n \rightarrow \mathcal{C}_n, \psi_n : \mathcal{C}_n \rightarrow \mathbb{A}^n)\}_{n=1}^\infty$ one has*

$$\mathbb{P}(\psi_n \circ \phi_n(X_1, \dots, X_n) \neq (X_1, \dots, X_n)) \geq 1 - \lambda.$$

Proof. Let $\mathcal{S}_n \subseteq \mathbb{A}^n$ denote the set of outcomes where $\psi_n \circ \phi_n(x_1, \dots, x_n) = (x_1, \dots, x_n)$, so that $|\mathcal{S}_n| \leq |\mathcal{C}_n| = m_n$. By the assumption that $\limsup_{n \rightarrow \infty} \frac{1}{n} \ln(m_n) < H(X_1)$, there exists $\delta \in (0, \lambda/2)$ and N_0 such that

$$\frac{1}{n} \ln(|\mathcal{S}_n|) \leq \frac{1}{n} \ln(m_n) < H(X_1) - 2\delta$$

for all $n > N_0$, and exponentiating this inequality gives

$$|\mathcal{S}_n| \leq |\mathcal{C}_n| = m_n < e^{n(H(X_1) - 2\delta)} \tag{1}$$

for all $n > N_0$. By Shannon-McMillan-Breiman's theorem (see last lecture), there exists an N_1 such that

$$\mathbb{P}((X_1, \dots, X_n) \notin A_\delta^n) < \delta \tag{2}$$

for all $n > N_1$. Therefore, by choosing

$$n > \max(N_0, N_1, \ln(\frac{2}{\lambda})/\delta),$$

we obtain the the following:

$$\begin{aligned} \mathbb{P}(\psi_n \circ \phi_n(X_1, \dots, X_n) = (X_1, \dots, X_n)) &= \sum_{x \in \mathcal{S}} \mathbb{P}_{X_1, \dots, X_n}(x_1, \dots, x_n) \\ &= \sum_{x \in \mathcal{S} \cap A_\delta^n} \mathbb{P}_{X_1, \dots, X_n}(x_1, \dots, x_n) + \sum_{x \in \mathcal{S} \cap (A_\delta^n)^c} \mathbb{P}_{X_1, \dots, X_n}(x_1, \dots, x_n) \\ &\leq |\mathcal{S}_n| \max_{x \in A_\delta^n} \mathbb{P}_{X_1, \dots, X_n}(x_1, \dots, x_n) + \sum_{x \in \mathcal{S} \cap (A_\delta^n)^c} \mathbb{P}_{X_1, \dots, X_n}(x_1, \dots, x_n) \\ &\left(\begin{array}{l} \text{the first term follows by (1)} \\ \text{and the definition of } A_\delta^n \end{array} \right) \leq e^{n(H(X_1) - 2\delta)} e^{-n(H(X_1) - \delta)} + \mathbb{P}_{X_1, \dots, X_n}(x \in (A_\delta^n)^c) \\ &\quad \left(\text{by (2)} \right) \leq e^{-n\delta} + \delta \\ &\quad \left(\begin{array}{l} \text{by } \delta \in (0, \frac{\lambda}{2}) \text{ and} \\ n > \ln(\frac{2}{\lambda})/\delta \end{array} \right) < \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda. \end{aligned}$$

Consequently, by considering the complement, we obtain

$$\mathbb{P}(\psi_n \circ \phi_n(X_1, \dots, X_n) \neq (X_1, \dots, X_n)) \geq 1 - \lambda.$$

□

As we see from the last two theorems, the value $\limsup_{n \rightarrow \infty} \frac{1}{n} \ln(m_n)$ plays a significant role in the theory of block coding. This justifies the following definition.

2.3.9 Definition. Given a block code sequence $\{(\mathcal{C}_n, \phi_n)\}_{n=1}^\infty$ with $m_n = |\mathcal{C}_n|$, we call

$$R = \limsup_{n \rightarrow \infty} \frac{1}{n} \ln(m_n)$$

the block code's compression rate.

In summary, if $R < H(X_1)$, then for all code sequences the probability of decoding error converges to 1 as $n \rightarrow \infty$. Conversely, if $R > H(X_1)$, then there exists a sequence of codes such that the probability of decoding error converges to 0 as $n \rightarrow \infty$.

2.3.10 Remark. Coding works because there are "small" $\{A_\epsilon^n\}$ of size $|A_\epsilon^n| \propto e^{nH(X_1)}$ with

$$\mathbb{P}_{X_1, \dots, X_n}(A_\epsilon^n) < \epsilon.$$

A natural followup goal is to find such sets for more general sources, which leads us to the next section.

2.4 Block Codes for Stationary Ergodic Sources

In the preceding section, we were able to provide results regarding how well reconstruction works under block coding for a DMS. In this section, we turn our attention to generalizing these results to other types processes, and it turns out that this can be done for a process which is stationary and ergodic. As we will see, much of this will work because we will be able to extend the AEP to this scenario.

2.4.11 Definition. A stationary source is a stochastic process $\{X_j\}_{j=-\infty}^{\infty}$ with the property that $\mathbb{P}(X \in A) = \mathbb{P}(X \circ \tau^j \in A)$, where $A = \{X_{i_1} = x_1, X_{i_2} = x_2, \dots, X_{i_n} = x_n\}$ and $(X \circ \tau^j)_l = X_{l+j}$ for all $l, j \in \mathbb{Z}$.

2.4.12 Definition. A source $\{X_j\}_{j=-\infty}^{\infty}$ is called ergodic if all events A such that $\{X \in A\} = \{X \circ \tau^j \in A\}$ satisfy $\mathbb{P}(X \in A) \in \{0, 1\}$.

In order to proceed, we state without proof Birkhoff's ergodic theorem, which will provide some of the necessary machinery for the rest of this section.

2.4.13 Theorem. (Birkhoff) *If $\{X_j\}_{j=-\infty}^{\infty}$ is stationary, then it is ergodic if and only if for every $k \in \mathbb{N}$ and for all functions $f : \mathbb{A}^k \rightarrow [0, \infty)$ such that $\mathbb{E}(f(X_{i_1}, \dots, X_{i_k})) < \infty$, we have for all $i_1, \dots, i_k \in \mathbb{Z}$ that*

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f(X_{i_1+j}, \dots, X_{i_k+j}) = \mathbb{E}(f(X_{i_1}, \dots, X_{i_k})) \right) = 1.$$