An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Aug. 1, 2016

Lecture #6B: Block to block coding of DMS

Outline of the lecture

- Consequence of Asymptotic Equipartition Property

Consequence of Asymptotic Equipartition Property Block to Block Coding of DMS

Outline of the lecture

- Consequence of Asymptotic Equipartition Property
- Block to block coding of DMS.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Outline

Asymptotic Equipartition Property

- Property 3 says that when L is large and ϵ is small, there are roughly $2^{LH(U)}\epsilon-$typical sequences $\mathbf{u}$.

Asymptotic Equipartition Property

- Property 3 says that when L is large and ϵ is small, there are roughly $2^{LH(U)}\epsilon-$typical sequences $\mathbf{u}$.
- Property 1 says each of these $\epsilon-$typical sequences has probability equal to $2^{-LH(U)}$.

Asymptotic Equipartition Property

- Property 3 says that when L is large and ϵ is small, there are roughly $2^{LH(U)}\epsilon-$typical sequences $\mathbf{u}$.
- Property 1 says each of these $\epsilon-$typical sequences has probability equal to $2^{-LH(U)}$.
- Property 2 says that the total probability of these $\epsilon-$ typical sequences is very nearly 1.

Asymptotic Equipartition Property

- Property 3 says that when L is large and ϵ is small, there are roughly $2^{LH(U)}\epsilon-$typical sequences $\mathbf{u}$.
- Property 1 says each of these $\epsilon-$typical sequences has probability equal to $2^{-LH(U)}$.
- Property 2 says that the total probability of these $\epsilon-$ typical sequences is very nearly 1.
- These three properties are known as *asymptotic equipartition property (AEP)* of the output sequence of a DMS.

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).
- We are interested in short description of these sequences.

Consequence of Asymptotic Equipartition Property Block to Block Coding of DMS

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).
- We are interested in short description of these sequences.
- Let us divide the set of sequences in X^n into typical set, $A_\epsilon^{(n)}$ and its complement.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).
- We are interested in short description of these sequences.
- Let us divide the set of sequences in X^n into typical set, $A_\epsilon^{(n)}$ and its complement.
- We will require $n(H+\epsilon)+1$ bits to represent the typical set and not more than $n\log|X|+1$ bits to represent its complement set.

Consequence of Asymptotic Equipartition Property Block to Block Coding of DMS

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).
- We are interested in short description of these sequences.
- Let us divide the set of sequences in X^n into typical set, $A_\epsilon^{(n)}$ and its complement.
- We will require $n(H+\epsilon)+1$ bits to represent the typical set and not more than $n\log|X|+1$ bits to represent its complement set.
- We can prefix the typical set by 0 and its complement by 1.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).
- We are interested in short description of these sequences.
- Let us divide the set of sequences in X^n into typical set, $A_\epsilon^{(n)}$ and its complement.
- We will require $n(H+\epsilon)+1$ bits to represent the typical set and not more than $n\log|X|+1$ bits to represent its complement set.
- We can prefix the typical set by 0 and its complement by 1.
- This code is one-to-one and easily decodable.

Consequences of Asymptotic Equipartition Property

- Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables drawn from the probability mass function p(x).
- We are interested in short description of these sequences.
- Let us divide the set of sequences in X^n into typical set, $A_\epsilon^{(n)}$ and its complement.
- We will require $n(H+\epsilon)+1$ bits to represent the typical set and not more than $n\log|X|+1$ bits to represent its complement set.
- We can prefix the typical set by 0 and its complement by 1.
- This code is one-to-one and easily decodable.
- Typical sequence requires short description of length $\approx nH$.

Consequences of Asymptotic Equipartition Property

- Expected codeword length is given by

$$\begin{aligned} E(l(X^n)) &= \sum_{x^n} p(x^n) l(x^n) \\ &= \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) l(x^n) + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n) l(x^n) \\ &\leq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n)[n(H+\epsilon)+2] + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n)[n \log |X| + 2] \\ &= Pr(A_\epsilon^{(n)})[n(H+\epsilon)+2] + Pr(A_\epsilon^{(n)c})[n \log |X| + 2] \\ &\leq n(H+\epsilon) + \epsilon[n \log |X|] + 2 \\ &= n(H+\epsilon') \end{aligned}$$

where $\epsilon' = \epsilon + \epsilon \log |X| + \frac{2}{n}$.

Outline

Block to Block Coding of DMS

- Given a K-ary DMS with output entropy $H(U)$ and given any positive numbers ϵ_1 and ϵ_2 , there exists, for all sufficiently large L, a D-ary block code of blocklength N for block message sets of blocklength L such that

$$\frac{N}{L} \leq \frac{H(U)}{\log D} + \epsilon_1$$

and such that the probability, P(F), that the codeword will not uniquely specify the message satisfies

$$P(F) < \epsilon_2$$

Proof:

Consequence of Asymptotic Equipartition Property Block to Block Coding of DMS

Block to Block Coding of DMS

- Given a K-ary DMS with output entropy $H(U)$ and given any positive numbers ϵ_1 and ϵ_2 , there exists, for all sufficiently large L, a D-ary block code of blocklength N for block message sets of blocklength L such that

$$\frac{N}{L} \leq \frac{H(U)}{\log D} + \epsilon_1$$

and such that the probability, P(F), that the codeword will not uniquely specify the message satisfies

$$P(F) < \epsilon_2$$

Proof:

- Suppose we assign a unique D-ary codeword of length N to each of the M ϵ-typical source output sequences $\mathbf{u}$, but use a single additional codeword to code all the non-typical source output sequences.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Block to Block Coding of DMS

- The smallest N that satisfies this condition is given by

$$D^{N-1} < M + 1 \leq D^N$$

Block to Block Coding of DMS

- The smallest N that satisfies this condition is given by

$$D^{N-1} < M + 1 \leq D^N$$

- Thus,

$$\begin{aligned} M &\geq D^{N-1} \\ or \quad (N-1)\log D &\leq \log M. \\ (N-1)\log D &\leq (1+\epsilon)LH(U) \\ or \quad \frac{N}{L} &\leq \frac{H(U)}{\log D} + \frac{\epsilon H(U)}{\log D} + \frac{1}{L} \end{aligned}$$

Block to Block Coding of DMS

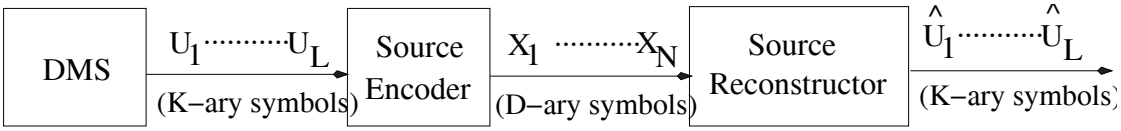
- The block-to-block (lossy) source coding theorem for a DMS states that given any positive numbers ϵ_1 and ϵ_2, there exists a source coding scheme as shown in Figure 1 for which

$$\frac{N}{L} < \frac{H(U)}{\log D} + \epsilon_1$$

and

$$P(F) < \epsilon_2$$

where P(F) is the probability that $[\hat{U}_1, \cdots, \hat{U}_L] \neq [U_1, \cdots, U_L]]$.



Block to Block Coding of DMS

- Defining the average error probability over the segment of L digits by

$$P_s = \frac{1}{L} \sum_{i=1}^{L} P_{ei}$$

and noting that $P_{ei} = P(\hat{U}_i \neq U_i) \leq P(F)$, we see that

$$P_s \leq P(F)$$

Block to Block Coding of DMS

- Defining the average error probability over the segment of L digits by

$$P_s = \frac{1}{L}\sum_{i=1}^{L} P_{ei}$$

and noting that $P_{ei} = P(\hat{U}_i \neq U_i) \leq P(F)$, we see that

$$P_s \leq P(F)$$

- Hence, we see that the lossy source coding theorem implies

$$P_s \leq \epsilon_2$$

so that P_s can be made arbitrarily small.

Block to Block Coding of DMS

- Show that P_s cannot be made arbitrarily small when N/L is smaller than $H(U)/\log D$. More precisely, if

$$\frac{N}{L} \leq \frac{H(U)}{\log D},$$

then

$$h(P_s) + P_s \log(K-1) \geq \left[\frac{H(U)}{\log D} - \frac{N}{L}\right] \log D$$

Proof:

Block to Block Coding of DMS

- Show that P_s cannot be made arbitrarily small when N/L is smaller than $H(U)/\log D$. More precisely, if

$$\frac{N}{L} \leq \frac{H(U)}{\log D},$$

then

$$h(P_s) + P_s \log(K-1) \geq \left[\frac{H(U)}{\log D} - \frac{N}{L}\right] \log D$$

Proof:

- We know that

$$\begin{aligned} I(U_1, \cdots, U_l; X_1 \cdots X_N) &= H(X_1 \cdots X_N) - H(X_1 \cdots X_N / U_1 \cdots U_L) \\ &\leq H(X_1 \cdots X_N) \end{aligned}$$

Block to Block Coding of DMS

- Using data processing lemma we have

$$\begin{aligned} I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) &\leq I(U_1 \cdots U_L; X_1 \cdots X_N) \\ &\leq H(X_1 \cdots X_N) \\ &\leq N \log D \end{aligned}$$

Block to Block Coding of DMS

- Using data processing lemma we have

$$\begin{aligned} I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) &\leq I(U_1 \cdots U_L; X_1 \cdots X_N) \\ &\leq H(X_1 \cdots X_N) \\ &\leq N \log D \end{aligned}$$

- Since $U_1 \cdots U_L$ are i.i.d.

$$H(U_1 \cdots U_L) = LH(U)$$

Block to Block Coding of DMS

- Using data processing lemma we have

$$\begin{aligned} I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) &\leq I(U_1 \cdots U_L; X_1 \cdots X_N) \\ &\leq H(X_1 \cdots X_N) \\ &\leq N \log D \end{aligned}$$

- Since $U_1 \cdots U_L$ are i.i.d.

$$H(U_1 \cdots U_L) = LH(U)$$

- We can write

$$\begin{aligned} H(U_1 \cdots U_L | \hat{U}_1 \cdots \hat{U}_L) &= H(U_1 \cdots U_L) - I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) \\ &\geq LH(U) - N \log D \end{aligned}$$

Block to Block Coding of DMS

- We know that

$$H(U_1 \cdots U_L | \hat{U}_1 \cdots \hat{U}_L) \leq \sum_{i=1}^{L} H(U_i | \hat{U}_i)$$

Block to Block Coding of DMS

- We know that

$$H(U_1 \cdots U_L | \hat{U}_1 \cdots \hat{U}_L) \leq \sum_{i=1}^{L} H(U_i | \hat{U}_i)$$

- Using Fano's lemma we have

$$\begin{aligned}
\sum_{i=1}^{L} H(U_i|\hat{U}_i) &\leq \sum_{i=1}^{L} \{H_2(P_{ei}) + P_{ei}\log_2(K-1)\} \\
&= \left\{\sum_{i=1}^{L} H_2(P_{ei})\right\} + LP_s\log_2(K-1) \\
\frac{1}{L}\sum_{i=1}^{L} H(U_i|\hat{U}_i) &\leq \left\{\frac{1}{L}\sum_{i=1}^{L} H_2(P_{ei})\right\} + P_s\log_2(K-1)
\end{aligned}$$

Block to Block Coding of DMS

- Since $H_2(p)$ is a concave function of p, we have

$$\frac{1}{L}\sum_{i=1}^{L} H_2(P_{ei}) \leq H_2\left(\frac{1}{L}\sum_{i=1}^{L} P_{ei}\right) = H_2(P_s)$$

Consequence of Asymptotic Equipartition Property
Block to Block Coding of DMS

Block to Block Coding of DMS

- Since $H_2(p)$ is a concave function of p, we have

$$\frac{1}{L}\sum_{i=1}^{L} H_2(P_{ei}) \leq H_2\left(\frac{1}{L}\sum_{i=1}^{L} P_{ei}\right) = H_2(P_s)$$

- Combining above equations we get

$$\frac{1}{L}H(U_1 \cdots U_L|\hat{U}_1 \cdots \hat{U}_L) \leq H_2(P_s) + P_s \log_2(K-1)$$

Block to Block Coding of DMS

- Since $H_2(p)$ is a concave function of p, we have

$$\frac{1}{L}\sum_{i=1}^{L} H_2(P_{ei}) \leq H_2\left(\frac{1}{L}\sum_{i=1}^{L} P_{ei}\right) = H_2(P_s)$$

- Combining above equations we get

$$\frac{1}{L}H(U_1 \cdots U_L|\hat{U}_1 \cdots \hat{U}_L) \leq H_2(P_s) + P_s \log_2(K-1)$$

- From above equations, we get

$$\begin{aligned} H_2(P_s) + P_s \log_2(K-1) &\geq \frac{1}{L}\{LH(U) - N\log D\} \\ &\geq \left\{\frac{H(U)}{\log D} - \frac{N}{L}\right\}\log D \end{aligned}$$