An introduction to Information Theory

Adrish Banerjee

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

Aug. 15, 2016

Lecture #8: Coding of sources with memory

Outline of the lecture

- Discrete stationery source (DSS)

Outline of the lecture

- Discrete stationery source (DSS)
- Block to variable length coding of DSS

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

Outline of the lecture

- Discrete stationery source (DSS)
- Block to variable length coding of DSS
- Coding of positive integers

Outline of the lecture

- Discrete stationery source (DSS)
- Block to variable length coding of DSS
- Coding of positive integers
- Elias-Willems Source Coding

Discrete stationery source

- A K-ary discrete stationary source (DSS) is a device that emits a sequence $U_i, \;\; i = 1, 2, \cdots$ of K-ary random variables such that for every $n \geq 1$ and every $L \geq 1$, the random vectors $[U_1, U_2, \cdots .U_L]$ and $[U_{n+1}, U_{n+2}, \cdots, U_{n+L}]$ have the same probability distribution.

Discrete stationery source

- A K-ary discrete stationary source (DSS) is a device that emits a sequence $U_i, \;\; i = 1, 2, \cdots$ of K-ary random variables such that for every $n \geq 1$ and every $L \geq 1$, the random vectors $[U_1, U_2, \cdots .U_L]$ and $[U_{n+1}, U_{n+2}, \cdots, U_{n+L}]$ have the same probability distribution.
- A DSS is said to have finite memory μ if, for all $n > \mu$,

  $$P_{U_n|U_1,\cdots,U_{n-1}}(u_n|u_1, \cdots, u_{n-1}) = P_{U_n|U_{n-\mu},\cdots,U_{n-1}}(u_n|u_{n-\mu}, \cdots, u_{n-1})$$

  holds for all choices of $u_1, \cdots, u_n$ and if μ is the smallest nonnegative integer such that this holds

Discrete stationery source

- A K-ary discrete stationary source (DSS) is a device that emits a sequence $U_i, \quad i = 1, 2, \cdots$ of K-ary random variables such that for every $n \geq 1$ and every $L \geq 1$, the random vectors $[U_1, U_2, \cdots .U_L]$ and $[U_{n+1}, U_{n+2}, \cdots, U_{n+L}]$ have the same probability distribution.
- A DSS is said to have finite memory μ if, for all $n > \mu$,

$$P_{U_n|U_1,\cdots,U_{n-1}}(u_n|u_1, \cdots, u_{n-1}) = P_{U_n|U_{n-\mu},\cdots,U_{n-1}}(u_n|u_{n-\mu}, \cdots, u_{n-1})$$

  holds for all choices of $u_1, \cdots, u_n$ and if μ is the smallest nonnegative integer such that this holds
- The discrete memory source (DMS) is a special case of DSS with memory $\mu = 0$.

Discrete stationery source

- For DSS with memory μ,

$$H(U_n|U_1, \cdots, U_{n-1}) = H(U_n|U_{n-\mu} \cdots, U_{n-1})$$

  for all $n > \mu$.

Discrete stationery source

- For DSS with memory μ,
$$H(U_n|U_1, \cdots, U_{n-1}) = H(U_n|U_{n-\mu} \cdots, U_{n-1})$$
for all $n > \mu$.
- Uncertainty per letter of output sequence of L letters is defined as
$$H_L(U) = \frac{1}{L} H(U_1, U_2, \cdots, U_L)$$

Discrete stationery source

- For DSS with memory μ,
$$H(U_n|U_1, \cdots, U_{n-1}) = H(U_n|U_{n-\mu} \cdots, U_{n-1})$$
for all $n > \mu$.
- Uncertainty per letter of output sequence of L letters is defined as
$$H_L(U) = \frac{1}{L} H(U_1, U_2, \cdots, U_L)$$
- Another way for defining uncertainty per letter of output sequence is $H(U_L|U_1U_2 \cdots U_{L-1})$

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then
- (i) $H(U_L|U_1 \cdots U_{L-1}) \leq H_L(U)$ for all $L \geq 1$.

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then
- (i) $H(U_L|U_1 \cdots U_{L-1}) \leq H_L(U)$ for all $L \geq 1$.
- (ii) $H(U_L|U_1 \cdots U_{L-1})$ is non-increasing as L increases

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then
- (i) $H(U_L|U_1 \cdots U_{L-1}) \leq H_L(U)$ for all $L \geq 1$.
- (ii) $H(U_L|U_1 \cdots U_{L-1})$ is non-increasing as L increases
- (iii) $H_L(U)$ is non-increasing as L increases, and

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then
- (i) $H(U_L|U_1 \cdots U_{L-1}) \leq H_L(U)$ for all $L \geq 1$.
- (ii) $H(U_L|U_1 \cdots U_{L-1})$ is non-increasing as L increases
- (iii) $H_L(U)$ is non-increasing as L increases, and
- (iv) $\lim_{L\to\infty} H(U_L|U_1 \cdots U_{L-1}) = \lim_{L\to\infty} H_L(U) = H_\infty(U)$

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then
- (i) $H(U_L|U_1 \cdots U_{L-1}) \leq H_L(U)$ for all $L \geq 1$.
- (ii) $H(U_L|U_1 \cdots U_{L-1})$ is non-increasing as L increases
- (iii) $H_L(U)$ is non-increasing as L increases, and
- (iv) $\lim_{L\to\infty} H(U_L|U_1 \cdots U_{L-1}) = \lim_{L\to\infty} H_L(U) = H_\infty(U)$

Proof:

- (ii) $H(U_{L+1}|U_1 \cdots U_L) \leq H(U_{L+1}|U_2 \cdots U_L) = H(U_L|U_1 \cdots U_{L-1})$

Discrete stationery source

- If $U_1, U_2, \cdots$ is the output sequence of a DSS, then
- (i) $H(U_L|U_1 \cdots U_{L-1}) \leq H_L(U)$ for all $L \geq 1$.
- (ii) $H(U_L|U_1 \cdots U_{L-1})$ is non-increasing as L increases
- (iii) $H_L(U)$ is non-increasing as L increases, and
- (iv) $\lim_{L\to\infty} H(U_L|U_1 \cdots U_{L-1}) = \lim_{L\to\infty} H_L(U) = H_\infty(U)$

Proof:

- (ii) $H(U_{L+1}|U_1 \cdots U_L) \leq H(U_{L+1}|U_2 \cdots U_L) = H(U_L|U_1 \cdots U_{L-1})$
- (i)

$$\begin{aligned} H(U_1U_2\cdots U_L) &= H(U_1) + H(U_2|U_1) + \cdots + H(U_L|U_1\cdots U_{L-1}) \\ &\geq LH(U_L|U_1\cdots U_{L-1}) \\ \implies H_L(U) &\geq H(U_L|U_1\cdots U_{L-1}) \end{aligned}$$

Discrete stationery source

Proof:

- (iii)

$$\begin{aligned} H(U_1U_2\cdots U_{L+1}) &= H(U_1\cdots U_L) + H(U_{L+1}|U_1\cdots U_L) \\ &\leq H(U_1\cdots U_L) + H(U_L|U_1\cdots U_{L-1}) \\ &\leq H(U_1\cdots U_L) + H_L(U) \\ &= LH_L(U) + H_L(U) \\ \implies \frac{H(U_1U_2\cdots U_{L+1})}{L+1} &\leq H_L(U) \\ H_{L+1}(U) &\leq H_L(U) \end{aligned}$$

Discrete stationery source

Proof:

- (iii)

$$\begin{aligned} H(U_1U_2\cdots U_{L+1}) &= H(U_1\cdots U_L)+H(U_{L+1}|U_1\cdots U_L)\\ &\leq H(U_1\cdots U_L)+H(U_L|U_1\cdots U_{L-1})\\ &\leq H(U_1\cdots U_L)+H_L(U)\\ &= LH_L(U)+H_L(U)\\ \implies \frac{H(U_1U_2\cdots U_{L+1})}{L+1} &\leq H_L(U)\\ H_{L+1}(U) &\leq H_L(U) \end{aligned}$$

- (iv) Since $H_L(U)\geq 0$ and $H(U_L|U_1\cdots U_{L-1})\geq 0$ and a non-increasing sequence bounded below always has a limit. Thus (ii) and (iii) ensure that both $\lim_{L\to\infty} H(U_L|U_1\cdots U_{L-1})$ and $\lim_{L\to\infty} H_L(U)$ always exist. (i) implies that

$$\lim_{L\to\infty} H(U_L|U_1\cdots U_{L-1}) \leq \lim_{L\to\infty} H_L(U) = H_\infty(U)$$

Discrete stationery source

Proof (contd.):

- Now we will have to show that equality holds. We know that

$$\begin{aligned} H(U_1\cdots U_L\cdots U_{L+n}) &= H(U_1\cdots U_L)+H(U_{L+1}|U_1\cdots U_L)+\cdots\\ &\quad +H(U_{L+n}|U_1\cdots U_{L+n-1})\\ &\leq H(U_1\cdots U_L)+nH(U_{L+1}|U_1\cdots U_L)\\ \frac{H(U_1\cdots U_L\cdots U_{L+n})}{n+L} &\leq \frac{H(U_1\cdots U_L)}{n+L}+\frac{n}{n+L}H(U_{L+1}|U_1\cdots U_L) \end{aligned}$$

Discrete stationery source

Proof (contd.):

- Now we will have to show that equality holds. We know that

$$\begin{aligned} H(U_1 \cdots U_L \cdots U_{L+n}) &= H(U_1 \cdots U_L) + H(U_{L+1}|U_1 \cdots U_L) + \cdots \\ &\quad + H(U_{L+n}|U_1 \cdots U_{L+n-1}) \\ &\leq H(U_1 \cdots U_L) + nH(U_{L+1}|U_1 \cdots U_L) \\ \frac{H(U_1 \cdots U_L \cdots U_{L+n})}{n+L} &\leq \frac{H(U_1 \cdots U_L)}{n+L} + \frac{n}{n+L} H(U_{L+1}|U_1 \cdots U_L) \end{aligned}$$

- Taking the limit as $n \to \infty$ on both sides we get,

$$H_\infty(U) \leq H(U_{L+1}|U_1 \cdots U_L)$$

Discrete stationery source

Proof (contd.):

- Now we will have to show that equality holds. We know that

$$\begin{aligned} H(U_1 \cdots U_L \cdots U_{L+n}) &= H(U_1 \cdots U_L) + H(U_{L+1}|U_1 \cdots U_L) + \cdots \\ &\quad + H(U_{L+n}|U_1 \cdots U_{L+n-1}) \\ &\leq H(U_1 \cdots U_L) + nH(U_{L+1}|U_1 \cdots U_L) \\ \frac{H(U_1 \cdots U_L \cdots U_{L+n})}{n+L} &\leq \frac{H(U_1 \cdots U_L)}{n+L} + \frac{n}{n+L} H(U_{L+1}|U_1 \cdots U_L) \end{aligned}$$

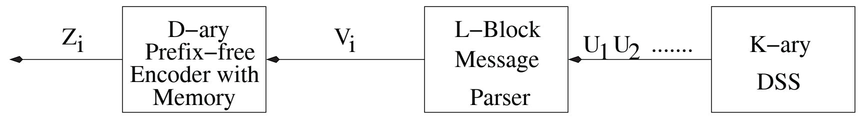
- Taking the limit as $n \to \infty$ on both sides we get,

$$H_\infty(U) \leq H(U_{L+1}|U_1 \cdots U_L)$$

- Now taking the limit as $L \to \infty$ on both sides we get,

$$H_\infty(U) \leq \lim_{L \to \infty} H(U_{L+1}|U_1 \cdots U_L)$$

Block to variable length coding of DSS

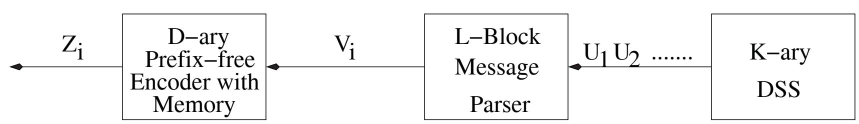


- If block length L messages from a DSS are encoded by a D-ary prefix-free code, where the code used for each message may depend on previous messages, then the length W_i of the codeword for the $i-$th message satisfies

$$\frac{E[W_i]}{L} \geq \frac{H_\infty(U)}{\log D}$$

where $H_\infty(U)$ is the uncertainty per letter of the DSS.

Block to variable length coding of DSS



- If block length L messages from a DSS are encoded by a D-ary prefix-free code, where the code used for each message may depend on previous messages, then the length W_i of the codeword for the $i-$th message satisfies

$$\frac{E[W_i]}{L} \geq \frac{H_\infty(U)}{\log D}$$

where $H_\infty(U)$ is the uncertainty per letter of the DSS.

Proof:

- Since we are using $L-$block message parser, the $i-$th message is encoded as

$$V_i = [U_{iL-L+1}, \cdots, U_{iL-1}, U_{iL}]$$

Block to variable length coding of DSS

- Given that $[V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}]$, it follows that the length W_i of the codeword Z_i for the message V_i must satisfy

$$E\left[W_i | [V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}]\right] \geq \frac{H(V_i | [V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}])}{\log D}$$

Block to variable length coding of DSS

- Given that $[V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}]$, it follows that the length W_i of the codeword Z_i for the message V_i must satisfy

$$E\left[W_i | [V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}]\right] \geq \frac{H(V_i | [V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}])}{\log D}$$

- Multiplying by $P_{V_1 \cdots V_{i-1}}(v_1, \cdots, v_{i-1})$ and summing over all values $v_1, \cdots, v_{i-1}$ gives

$$E\left[W_i\right] \geq \frac{H(V_i | [V_1, \cdots, V_{i-1}])}{\log D}$$

Block to variable length coding of DSS

- Given that $[V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}]$, it follows that the length W_i of the codeword Z_i for the message V_i must satisfy

$$E[W_i | [V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}]] \geq \frac{H(V_i | [V_1, \cdots, V_{i-1}] = [v_1, \cdots, v_{i-1}])}{\log D}$$

- Multiplying by $P_{V_1 \cdots V_{i-1}}(v_1, \cdots, v_{i-1})$ and summing over all values $v_1, \cdots, v_{i-1}$ gives

$$E[W_i] \geq \frac{H(V_i | [V_1, \cdots, V_{i-1}])}{\log D}$$

- We can write

$$\begin{aligned} H(V_i | V_1 \cdots V_{i-1}) &= H(U_{iL-L+1} \cdots U_{iL} | U_1 \cdots U_{iL-L}) \\ &= H(U_{iL-L+1} | U_1 \cdots U_{iL-L}) + \cdots + \\ &\quad H(U_{iL} | U_1 \cdots U_{iL-1}) \\ &\geq LH(U_{iL} | U_1 \cdots U_{iL-1}) \\ &\geq LH_\infty(U) \end{aligned}$$

Coding of positive integers

- Binary representation

n	B(n)
1	1
2	1 0
3	1 1
4	1 0 0
5	1 0 1
6	1 1 0
7	1 1 1
8	1 0 0 0

Coding of positive integers

- Binary representation

n	B(n)
1	1
2	1 0
3	1 1
4	1 0 0
5	1 0 1
6	1 1 0
7	1 1 1
8	1 0 0 0

- This is not a prefix-free code.

Coding of positive integers

- Binary representation

n	B(n)
1	1
2	1 0
3	1 1
4	1 0 0
5	1 0 1
6	1 1 0
7	1 1 1
8	1 0 0 0

- This is not a prefix-free code.
- Every codeword as 1 as the first digit.

Coding of positive integers

- Binary representation

n	B(n)
1	1
2	1 0
3	1 1
4	1 0 0
5	1 0 1
6	1 1 0
7	1 1 1
8	1 0 0 0

- This is not a prefix-free code.
- Every codeword as 1 as the first digit.
- The length of the codeword $L(n) = \lfloor \log_2 n \rfloor + 1$.

Coding of positive integers

- Elias code 1

n	$B_1(n)$
1	1
2	0 1 0
3	0 1 1
4	0 0 1 0 0
5	0 0 1 0 1
6	0 0 1 1 0
7	0 0 1 1 1
8	0 0 0 1 0 0 0

Coding of positive integers

- Elias code 1

n	$B_1(n)$
1	1
2	0 1 0
3	0 1 1
4	0 0 1 0 0
5	0 0 1 0 1
6	0 0 1 1 0
7	0 0 1 1 1
8	0 0 0 1 0 0 0

- This is a prefix-free code.

Coding of positive integers

- Elias code 1

n	$B_1(n)$
1	1
2	0 1 0
3	0 1 1
4	0 0 1 0 0
5	0 0 1 0 1
6	0 0 1 1 0
7	0 0 1 1 1
8	0 0 0 1 0 0 0

- This is a prefix-free code.
- No of leading zeros tell how many digits follow the leading 1.

Coding of positive integers

- Elias code 1

n	$B_1(n)$
1	1
2	0 1 0
3	0 1 1
4	0 0 1 0 0
5	0 0 1 0 1
6	0 0 1 1 0
7	0 0 1 1 1
8	0 0 0 1 0 0 0

- This is a prefix-free code.
- No of leading zeros tell how many digits follow the leading 1.
- The length of the codeword $L_1(n) = 2\lfloor \log_2 n \rfloor + 1$.

Coding of positive integers

- Elias code 2

n	$B_2(n)$
1	1
2	0 1 0 0
3	0 1 0 1
4	0 1 1 0 0
5	0 1 1 0 1
6	0 1 1 1 0
7	0 1 1 1 1
8	0 0 1 0 0 0 0 0

Coding of positive integers

- Elias code 2

n	$B_2(n)$
1	1
2	0 1 0 0
3	0 1 0 1
4	0 1 1 0 0
5	0 1 1 0 1
6	0 1 1 1 0
7	0 1 1 1 1
8	0 0 1 0 0 0 0 0

- Elias code 2 begins with $B_1(L(n))$ (Elias code 1 to signal the number of following digits in the codeword) followed by B(n) with leading 1 removed.

Coding of positive integers

- Elias code 2

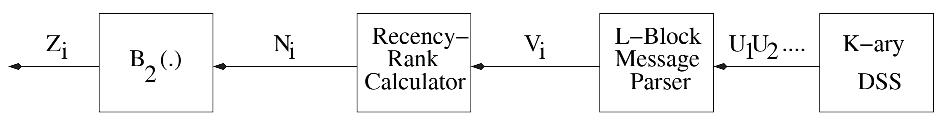
Coding of positive integers

- Elias code 2
- The length of the codeword $L_2(n)$ is given by

$$\begin{aligned} L_2(n) &= L_1(L(n)) + L(n) - 1 \\ &= L_1(\lfloor \log_2 n \rfloor + 1) + \lfloor \log_2 n \rfloor \\ &= \lfloor \log_2 n \rfloor + 2\lfloor \log_2(\lfloor \log_2 n \rfloor + 1) \rfloor + 1 \end{aligned}$$

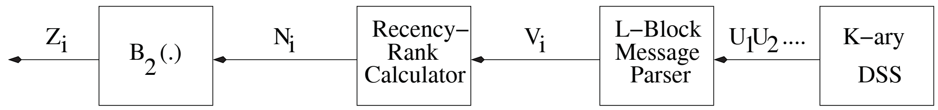
.

Elias-Willems Source Coding



- The message seen most recently has recency rank of 1, next different message value most recently seen has recency rank 2 and so on.

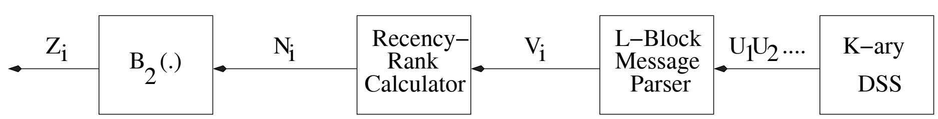
Elias-Willems Source Coding



- The message seen most recently has recency rank of 1, next different message value most recently seen has recency rank 2 and so on.
- Consider $L = 2$, and $K = 2$, i.e. four possible message values are $00, 01, 10, 11$. Consider the following sequence of past messages

$$\cdots|V_{i-2}|V_{i-1} = \cdots|11|01|10|01|01|00|01$$

Elias-Willems Source Coding



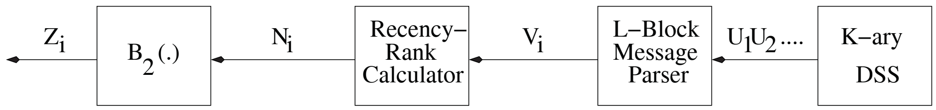
- The message seen most recently has recency rank of 1, next different message value most recently seen has recency rank 2 and so on.
- Consider $L = 2$, and $K = 2$, i.e. four possible message values are $00, 01, 10, 11$. Consider the following sequence of past messages

$$\cdots|V_{i-2}|V_{i-1} = \cdots|11|01|10|01|01|00|01$$

- In this example, message 01 has a recency rank of 1, message 00 has a recency rank of 2, message 10 has a recency rank of 3, and message 11 has a recency rank of 4.

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

Elias-Willems Source Coding



- The message seen most recently has recency rank of 1, next different message value most recently seen has recency rank 2 and so on.
- Consider $L = 2$, and $K = 2$, i.e. four possible message values are $00, 01, 10, 11$. Consider the following sequence of past messages

$$\cdots|V_{i-2}|V_{i-1} = \cdots|11|01|10|01|01|00|01$$

- In this example, message 01 has a recency rank of 1, message 00 has a recency rank of 2, message 10 has a recency rank of 3, and message 11 has a recency rank of 4.
- In Elias Willems source coding scheme, recency rank calculator assigns a recency rank, N_i to the message V_i.

Elias-Willems Source Coding

- For the Elias-Willems binary coding scheme for block length L messages from a K-ary ergodic discrete stationary source, the codeword length W_i for all i sufficiently large satisfies

$$\frac{E[W_i]}{L} \leq H_L(U) + \frac{2}{L}\log_2(LH_L(U)+1) + \frac{1}{L}$$

Elias-Willems Source Coding

- For the Elias-Willems binary coding scheme for block length L messages from a K-ary ergodic discrete stationary source, the codeword length W_i for all i sufficiently large satisfies

$$\frac{E[W_i]}{L} \leq H_L(U) + \frac{2}{L}\log_2(LH_L(U)+1) + \frac{1}{L}$$

Proof:

- We define time Δ_i to the recent occurence of message V_i as $\Delta_i = \delta$, where if $V_i = v$, $V_{i-\delta}$ is the most recent past message with the same value of v. Also $N_i \leq \Delta_i$.

Elias-Willems Source Coding

- For the Elias-Willems binary coding scheme for block length L messages from a K-ary ergodic discrete stationary source, the codeword length W_i for all i sufficiently large satisfies

$$\frac{E[W_i]}{L} \leq H_L(U) + \frac{2}{L}\log_2(LH_L(U)+1) + \frac{1}{L}$$

Proof:

- We define time Δ_i to the recent occurence of message V_i as $\Delta_i = \delta$, where if $V_i = v$, $V_{i-\delta}$ is the most recent past message with the same value of v. Also $N_i \leq \Delta_i$.
- We can write

$$E[\Delta_i | V_i = v] = \frac{1}{P_V(v)}$$

Elias-Willems Source Coding

- Codeword length W_i can be upper bounded by

$$\begin{aligned} W_i = L_2(N_i) &\leq L_2(\Delta_i) \\ &\leq \log_2(\Delta_i) + 2\log_2(\log_2(\Delta_i)+1)+1 \end{aligned}$$

Elias-Willems Source Coding

- Codeword length W_i can be upper bounded by

$$\begin{aligned} W_i = L_2(N_i) &\leq L_2(\Delta_i) \\ &\leq \log_2(\Delta_i) + 2\log_2(\log_2(\Delta_i)+1)+1 \end{aligned}$$

- Taking the conditional expectation, given that $V_i = v$ and applying Jensen's inequality, we get

$$E[W_i|V_i = v] \leq \log_2(E[\Delta_i|V_i = v]) + 2\log_2(\log_2(E[\Delta_i|V_i = v]) + 1) + 1$$

Elias-Willems Source Coding

- Codeword length W_i can be upper bounded by

$$W_i = L_2(N_i) \leq L_2(\Delta_i) \leq \log_2(\Delta_i) + 2\log_2(\log_2(\Delta_i)+1)+1$$

- Taking the conditional expectation, given that $V_i = v$ and applying Jensen's inequality, we get

$$E[W_i|V_i = v] \leq \log_2(E[\Delta_i|V_i = v]) + 2\log_2(\log_2(E[\Delta_i|V_i = v])+1)+1$$

- Substituting the value of $E[\Delta_i|V_i = v]$ in the above expression, we get

$$E[W_i|V_i = v] \leq -\log_2(P_V(v)) + 2\log_2(-\log_2(P_V(v))+1)+1$$

Elias-Willems Source Coding

- Codeword length W_i can be upper bounded by

$$W_i = L_2(N_i) \leq L_2(\Delta_i) \leq \log_2(\Delta_i) + 2\log_2(\log_2(\Delta_i)+1)+1$$

- Taking the conditional expectation, given that $V_i = v$ and applying Jensen's inequality, we get

$$E[W_i|V_i = v] \leq \log_2(E[\Delta_i|V_i = v]) + 2\log_2(\log_2(E[\Delta_i|V_i = v])+1)+1$$

- Substituting the value of $E[\Delta_i|V_i = v]$ in the above expression, we get

$$E[W_i|V_i = v] \leq -\log_2(P_V(v)) + 2\log_2(-\log_2(P_V(v))+1)+1$$

- Multiplying the above expression by $P_V(v)$, summing over v and using Jensen's inequality, we get

$$E[W_i] \leq H(V) + 2\log_2(H(V)+1)+1$$

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

Elias-Willems Source Coding

- Also

$$\begin{aligned} H(V) &= H(U_1 U_2 \cdots U_L) \\ &= L H_L(U). \end{aligned}$$

Elias-Willems Source Coding

- Also

$$\begin{aligned} H(V) &= H(U_1 U_2 \cdots U_L) \\ &= L H_L(U). \end{aligned}$$

- Substituting this value of $H(V)$ in the previous expression and dividing by L, we get

$$\frac{E[W_i]}{L} \leq H_L(U) + \frac{2}{L}\log_2(LH_L(U) + 1) + \frac{1}{L}$$