An introduction to Information Theory

Adrish Banerjee

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

Aug. 15, 2016

Lecture #9: Channel capacity

Outline of the lecture

- Discrete Memoryless Channel (DMC)

Outline of the lecture

- Discrete Memoryless Channel (DMC)
- Uniformly dispersive channel

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 Memoryless Channel (DMC)
- Uniformly dispersive channel
- Uniformly focusing channel

Outline of the lecture

- Discrete Memoryless Channel (DMC)
- Uniformly dispersive channel
- Uniformly focusing channel
- Strongly symmetric channel

Outline of the lecture

- Discrete Memoryless Channel (DMC)
- Uniformly dispersive channel
- Uniformly focusing channel
- Strongly symmetric channel
- Capacity of strongly symmetric channel

Outline of the lecture

- Discrete Memoryless Channel (DMC)
- Uniformly dispersive channel
- Uniformly focusing channel
- Strongly symmetric channel
- Capacity of strongly symmetric channel
- Symmetric channel

Outline of the lecture

- Discrete Memoryless Channel (DMC)
- Uniformly dispersive channel
- Uniformly focusing channel
- Strongly symmetric channel
- Capacity of strongly symmetric channel
- Symmetric channel
- Capacity of symmetric channel

Discrete Memoryless Channel

- A discrete memoryless channel (DMC) is specified by following three quantities:

Discrete Memoryless Channel

- A discrete memoryless channel (DMC) is specified by following three quantities:
  - An input alphabet, A.

Discrete Memoryless Channel

- A discrete memoryless channel (DMC) is specified by following three quantities:
  - An input alphabet, A.
  - An output alphabet, B.

Discrete Memoryless Channel

- A discrete memoryless channel (DMC) is specified by following three quantities:
  - An input alphabet, A.
  - An output alphabet, B.
  - The conditional probability distribution $P_{Y|X}(\cdot|x)$ over B for $x \in$ A such that

$$P(y_n|x_1, \cdots, x_{n-1}, x_n, y_1, \cdots, y_{n-1}) = P_{Y|X}(y_n|x_n)$$

Discrete Memoryless Channel

- A discrete memoryless channel (DMC) is specified by following three quantities:
  - An input alphabet, A.
  - An output alphabet, B.
  - The conditional probability distribution $P_{Y|X}(\cdot|x)$ over B for $x \in$ A such that

$$P(y_n|x_1, \cdots, x_{n-1}, x_n, y_1, \cdots, y_{n-1}) = P_{Y|X}(y_n|x_n)$$

- DMC without feedback is described by

$$P(x_n|x_1, \cdots, x_{n-1}, y_1, \cdots, y_{n-1}) = P(x_n|x_1, \cdots, x_{n-1})$$

Discrete Memoryless Channel

- For a DMC without feedback,

$$P(y_1, \cdots, y_n | x_1, \cdots, x_n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \text{ for } n = 1, 2, \cdots$$

Discrete Memoryless Channel

- For a DMC without feedback,

$$P(y_1, \cdots, y_n | x_1, \cdots, x_n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \text{ for } n = 1, 2, \cdots$$

- Proof:

$$\begin{aligned}
P(x_1, \cdot\cdot, x_n, y_1, \cdot\cdot, y_n) &= \prod_{i=1}^{n} P(x_i | x_1, \cdot\cdot, x_{i-1}, y_1, \cdot\cdot, y_{i-1}) P(y_i | x_1, \cdot\cdot, x_i, y_1, \cdot\cdot, y_{i-1}) \\
&= \prod_{i=1}^{n} P(x_i | x_1, \cdots, x_{i-1}) P_{Y|X}(y_i|x_i) \\
&= \left[\prod_{j=1}^{n} P(x_j | x_1, \cdots, x_{j-1}) \right] \left[\prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \right] \\
&= P(x_1, \cdots, x_n) \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)
\end{aligned}$$

Discrete Memoryless Channel

- For a DMC without feedback,

$$P(y_1, \cdots, y_n | x_1, \cdots, x_n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \text{ for } n = 1, 2, \cdots$$

- Proof:

$$\begin{aligned} P(x_1, \cdot\cdot, x_n, y_1, \cdot\cdot, y_n) &= \prod_{i=1}^{n} P(x_i|x_1, \cdot\cdot, x_{i-1}, y_1, \cdot\cdot, y_{i-1}) P(y_i|x_1, \cdot\cdot, x_i, y_1, \cdot\cdot, y_{i-1}) \\ &= \prod_{i=1}^{n} P(x_i|x_1, \cdots, x_{i-1}) P_{Y|X}(y_i|x_i) \\ &= \left[\prod_{j=1}^{n} P(x_j|x_1, \cdots, x_{j-1})\right] \left[\prod_{i=1}^{n} P_{Y|X}(y_i|x_i)\right] \\ &= P(x_1, \cdots, x_n) \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \end{aligned}$$

  - Dividing both sides by $P(x_1, \cdots, x_n)$, we get the desired expression.

Channel capacity

- Channel capacity of a DMC is defined as maximum average mutual information $I(X; Y)$ that can be obtained by the choice of $P(x)$, i.e.

$$C = \max_{P_X} I(X; Y)$$

Channel capacity

- Channel capacity of a DMC is defined as maximum average mutual information $I(X;Y)$ that can be obtained by the choice of $P(x)$, i.e.

$$C = \max_{P_X} I(X;Y)$$

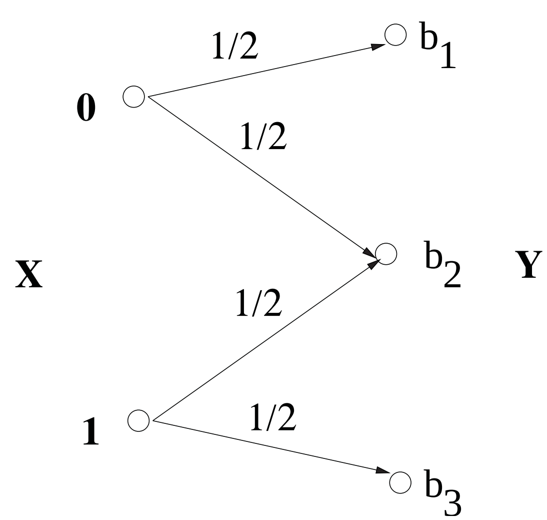
- Equivalently

$$C = \max_{P_X} [H(Y) - H(Y|X)]$$

Uniformly dispersive channel

- Let DMC has K inputs and J outputs. We say a DMC is uniformly dispersive if the probabilities of the J transitions leaving an input when put in decreasing order has same values for all K inputs.

Example:



Uniformly dispersive channel

- Independent of choice of P_X, for uniformly dispersive channel

$$H(Y|X) = -\sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_j$ are transition probabilities leaving each input letter.

Uniformly dispersive channel

- Independent of choice of P_X, for uniformly dispersive channel

$$H(Y|X) = -\sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_j$ are transition probabilities leaving each input letter.

Proof:

- From the definition of a uniformly dispersive channel, we know that

$$H(Y|X = a_k) = -\sum_{j=1}^{J} p_j \log p_j \text{ for } k = 1, \cdots, K.$$

Uniformly dispersive channel

- Independent of choice of P_X, for uniformly dispersive channel

$$H(Y|X) = -\sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_J$ are transition probabilities leaving each input letter.

Proof:

- From the definition of a uniformly dispersive channel, we know that

$$H(Y|X = a_k) = -\sum_{j=1}^{J} p_j \log p_j \quad \text{for} \quad k = 1, \cdots, K.$$

- From the definition of $H(Y|X)$, it follows then

$$H(Y|X) = \sum_{k=1}^{K} P(a_k) H(Y/X = a_k) = H(Y|X = a_k)$$

Uniformly dispersive channel

- Capacity for a uniformly dispersive channel is given by

$$C = \max_{P_X} [H(Y)] + \sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_J$ are transition probabilities leaving each input letter.

Uniformly dispersive channel

- Capacity for a uniformly dispersive channel is given by

$$C = \max_{P_X} [H(Y)] + \sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_j$ are transition probabilities leaving each input letter.

Proof:

- Since for a uniformly dispersive channel,

$$H(Y|X) = -\sum_{j=1}^{J} p_j \log p_j$$

Uniformly dispersive channel

- Capacity for a uniformly dispersive channel is given by

$$C = \max_{P_X} [H(Y)] + \sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_j$ are transition probabilities leaving each input letter.

Proof:

- Since for a uniformly dispersive channel,

$$H(Y|X) = -\sum_{j=1}^{J} p_j \log p_j$$

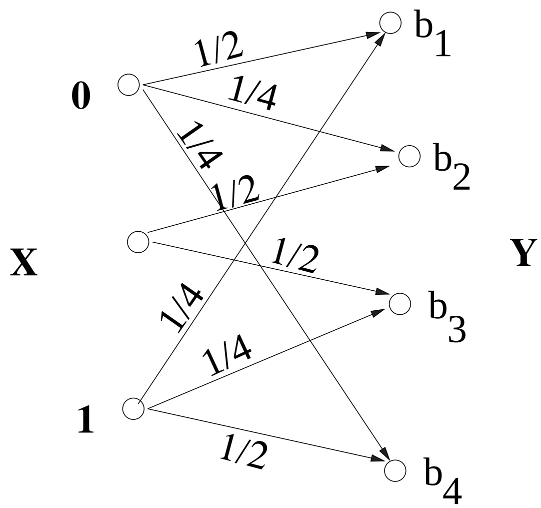
- Therefore from the definition of channel capacity, we get

$$\begin{aligned} C &= \max_{P_X} [H(Y) - H(Y|X)] \\ &= \max_{P_X} [H(Y)] + \sum_{j=1}^{J} p_j \log p_j \end{aligned}$$

Uniformly focusing channel

- Let DMC has K inputs and J outputs. We say a DMC is uniformly focusing if the probabilities of the K transitions to an output letter when put in decreasing order has same values for all J outputs.

Example:



Uniformly focusing channel

- For a K-input, J-output uniformly focusing channel, uniform input probability distribution will result in uniform output probability distribution.

Uniformly focusing channel

- For a K-input, J-output uniformly focusing channel, uniform input probability distribution will result in uniform output probability distribution.
- Also,

$$\max_{P_X} [H(Y)] = \log J$$

Uniformly focusing channel

- For a K-input, J-output uniformly focusing channel, uniform input probability distribution will result in uniform output probability distribution.
- Also,

$$\max_{P_X} [H(Y)] = \log J$$

Proof:

$$P(y) = \sum_x P(y|x)P(x) = \frac{1}{K}\sum_x P(y|x)$$

- Since DMC is uniformly focusing channel, sum of the right hand side will be same for all J values of y.

Uniformly focusing channel

- For a K-input, J-output uniformly focusing channel, uniform input probability distribution will result in uniform output probability distribution.
- Also,

$$\max_{P_X}[H(Y)] = \log J$$

Proof:

$$P(y) = \sum_x P(y|x)P(x) = \frac{1}{K}\sum_x P(y|x)$$

- Since DMC is uniformly focusing channel, sum of the right hand side will be same for all J values of y.
- Hence $P(y)$ is uniformly distributed and it follows from the property of entropy that

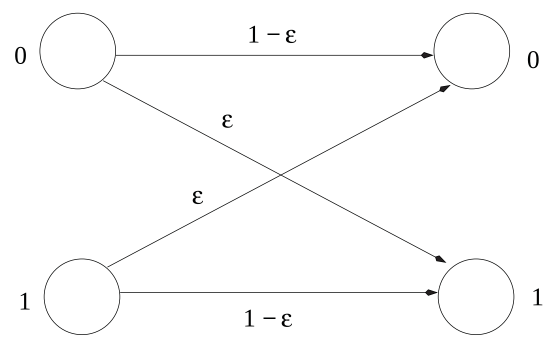
$$\max_{P_X}[H(Y)] = \log J$$

and is achieved by uniform input distribution.

Strongly symmetric channel

- DMC is strongly symmetric if it is both uniformly dispersive and uniformly focusing.

Example:



Strongly symmetric channel

- For a strongly symmetric channel with K inputs and J outputs, the channel capacity is given by

$$C = \log J + \sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_J$ are the transition probabilities leaving an input letter.

Strongly symmetric channel

- For a strongly symmetric channel with K inputs and J outputs, the channel capacity is given by

$$C = \log J + \sum_{j=1}^{J} p_j \log p_j$$

where $p_1, p_2, \cdots, p_J$ are the transition probabilities leaving an input letter.

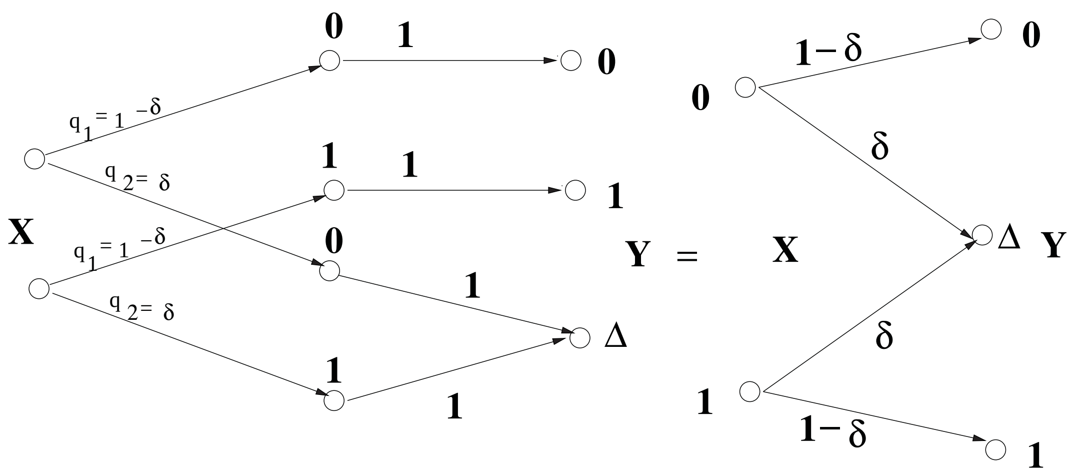
- Also, the uniform input distribution

$$P(x) = \frac{1}{K} \quad \text{all } x$$

achieves capacity.

Symmetric channel

- A DMC is said to be symmetric if it can be decomposed into L strongly symmetric channels with selection probabilities $q_1, q_2, \cdots, q_L$.



Symmetric channel

Algorithm to determine if the DMC is symmetric

1. Partition the output letters into subsets $B^{(1)}, B^{(2)}, \cdots, B^{(L)}$ such that two output letters are in the same subset if and only if they have the "same focusing". Set $i = 1$.

Symmetric channel

Algorithm to determine if the DMC is symmetric

1. Partition the output letters into subsets $B^{(1)}, B^{(2)}, \cdots, B^{(L)}$ such that two output letters are in the same subset if and only if they have the "same focusing". Set $i = 1$.
2. Check if all input letters have the "same dispersion" into the subset $B^{(i)}$ of output letters. If yes, set q_i equal to the sum of probabilities into $B^{(i)}$ from any input letter.

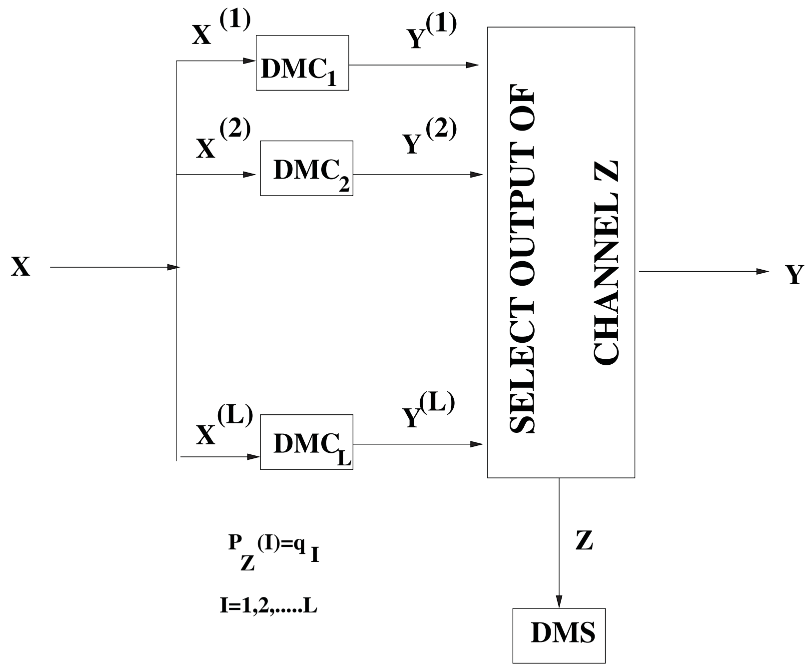
Symmetric channel

Algorithm to determine if the DMC is symmetric

1. Partition the output letters into subsets $B^{(1)}, B^{(2)}, \cdots, B^{(L)}$ such that two output letters are in the same subset if and only if they have the "same focusing". Set $i = 1$.
2. Check if all input letters have the "same dispersion" into the subset $B^{(i)}$ of output letters. If yes, set q_i equal to the sum of probabilities into $B^{(i)}$ from any input letter.
3. If $i = L$ stop. The channel is symmetric and the selection probabilities $q_1, q_2, \cdots, q_L$ have been found. If $i < L$, increment i and return to previous step.

Symmetric channel

General DMC created by L DMC's with selection probabilities $q_1, \cdots, q_L$.



Symmetric channel

General DMC created by L DMC's with selection probabilities $q_1, \cdots, q_L$.
Assumptions:

- All component DMC's have the same alphabet, i.e. $A = A^{(1)} = A^{(2)} = \cdots = A^{(L)}$.

Symmetric channel

General DMC created by L DMC's with selection probabilities $q_1, \cdots, q_L$.
Assumptions:

- All component DMC's have the same alphabet, i.e. $A = A^{(1)} = A^{(2)} = \cdots = A^{(L)}$.
- All component DMC's have disjoint output alphabets, i.e.

$$B^{(i)} \cap B^{(j)} = \phi, \; i \neq j \; \text{ and } \; B = B^{(1)} \cup B^{(2)} \cup \cdots \cup B^{(L)}$$

Symmetric channel

General DMC created by L DMC's with selection probabilities $q_1, \cdots, q_L$.
Assumptions:

- All component DMC's have the same alphabet, i.e. $A = A^{(1)} = A^{(2)} = \cdots = A^{(L)}$.
- All component DMC's have disjoint output alphabets, i.e.

$$B^{(i)} \cap B^{(j)} = \phi, \; i \neq j \; \text{ and } \; B = B^{(1)} \cup B^{(2)} \cup \cdots \cup B^{(L)}$$

- X and Z are statistically independent.

Symmetric channel

General DMC created by L DMC's with selection probabilities $q_1, \cdots, q_L$.
Assumptions:

- All component DMC's have the same alphabet, i.e. $A = A^{(1)} = A^{(2)} = \cdots = A^{(L)}$.
- All component DMC's have disjoint output alphabets, i.e.

$$B^{(i)} \cap B^{(j)} = \phi, \; i \neq j \text{ and } B = B^{(1)} \cup B^{(2)} \cup \cdots \cup B^{(L)}$$

- X and Z are statistically independent.
- $P_Z(i) = q_i$.

Symmetric channel

General DMC created by L DMC's with selection probabilities $q_1, \cdots, q_L$.
Assumptions:

- All component DMC's have the same alphabet, i.e. $A = A^{(1)} = A^{(2)} = \cdots = A^{(L)}$.
- All component DMC's have disjoint output alphabets, i.e.

$$B^{(i)} \cap B^{(j)} = \phi, \; i \neq j \text{ and } B = B^{(1)} \cup B^{(2)} \cup \cdots \cup B^{(L)}$$

- X and Z are statistically independent.
- $P_Z(i) = q_i$.

Claim: For this general DMC,

$$I(X;Y) = \sum_{i=1}^{L} q_i I(X; Y^{(i)})$$

Symmetric channel

- Proof of Claim:

$$\begin{aligned} H(YZ) &= H(Y) + H(Z|Y) = H(Y) \\ &= H(Z) + H(Y|Z) \end{aligned}$$

Symmetric channel

- Proof of Claim:

$$\begin{aligned} H(YZ) &= H(Y) + H(Z|Y) = H(Y) \\ &= H(Z) + H(Y|Z) \end{aligned}$$

- Thus

$$\begin{aligned} H(Y) &= H(Z) + H(Y|Z) \\ &= H(Z) + \sum_{i=1}^{L} H(Y|Z=i)P_Z(i) \\ &= H(Z) + \sum_{i=1}^{L} H(Y^{(i)})q_i \end{aligned}$$

Symmetric channel

- Similarly

$$
\begin{aligned}
H(YZ|X) &= H(Y|X) + H(Z|XY) = H(Y|X) \\
&= H(Z|X) + H(Y|XZ) = H(Z) + H(Y|XZ)
\end{aligned}
$$

Symmetric channel

- Thus,

$$
\begin{aligned}
H(Y|X) &= H(Z) + H(Y|XZ) \\
&= H(Z) + \sum_{i=1}^{L} H(Y|X, Z = i) P_Z(i) \\
&= H(Z) + \sum_{i=1}^{L} H(Y^{(i)}|X) q_i
\end{aligned}
$$

Symmetric channel

- Therefore

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= \sum_{i=1}^{L} q_i \left[H(Y^{(i)}) - H(Y^{(i)}|X) \right] \\
&= \sum_{i=1}^{L} q_i I(X;Y^{(i)})
\end{aligned}
$$

Symmetric channel

- For a symmetric DMC,

$$C = \sum_{i=1}^{L} q_i C_i$$

where C_i is the capacity of the $i-$th strongly symmetric DMC and q_i is its selection probability. Also uniform input distribution achieves capacity.

Symmetric channel

- For a symmetric DMC,

$$C = \sum_{i=1}^{L} q_i C_i$$

where C_i is the capacity of the $i-$th strongly symmetric DMC and q_i is its selection probability. Also uniform input distribution achieves capacity.

Proof:

- We have just shown (in the previous slide) that

$$I(X;Y) = \sum_{i=1}^{L} q_i I(X;Y^{(i)})$$

Symmetric channel

- For a symmetric DMC,

$$C = \sum_{i=1}^{L} q_i C_i$$

where C_i is the capacity of the $i-$th strongly symmetric DMC and q_i is its selection probability. Also uniform input distribution achieves capacity.

Proof:

- We have just shown (in the previous slide) that

$$I(X;Y) = \sum_{i=1}^{L} q_i I(X;Y^{(i)})$$

- Also

$$C = \max_{P_X} I(X;Y) \leq \sum_{i=1}^{L} q_i \max_{P_X} I(X;Y^{(i)}) = \sum_{i=1}^{L} q_i C_i$$

Symmetric channel

- Above expression is satisfied with equality if and only if there is a P_X that simultaneously maximizes $I(X;Y^{(1)}), I(X;Y^{(2)}), \cdots, I(X;Y^{(L)})$

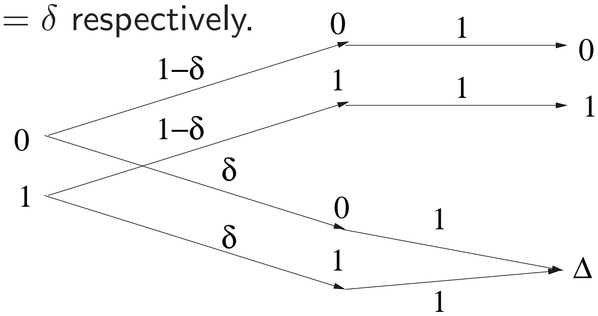
Symmetric channel

- Above expression is satisfied with equality if and only if there is a P_X that simultaneously maximizes $I(X;Y^{(1)}), I(X;Y^{(2)}), \cdots, I(X;Y^{(L)})$
- Since each of the component DMC is strongly symmetric channel, uniform input probability distribution, i.e. $P_X(a_k) = 1/K \ \forall k$ is one such P_X.

Symmetric channel

Capacity.of Binary Erasure Channel (BEC)

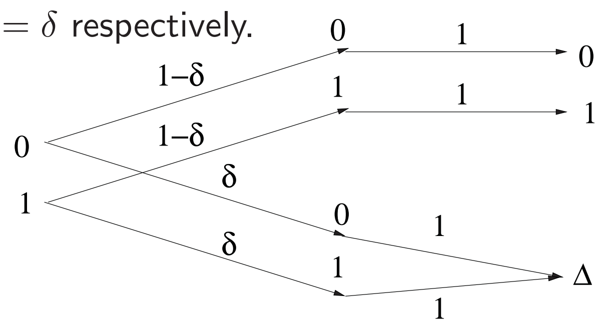
- BEC can be decomposed into $L = 2$ strongly symmetric channels with capacities $C_1 = 1$, $C_2 = 0$ and selection probabilities $q_1 = 1 - \delta$, $q_2 = \delta$ respectively.



Symmetric channel

Capacity.of Binary Erasure Channel (BEC)

- BEC can be decomposed into $L = 2$ strongly symmetric channels with capacities $C_1 = 1$, $C_2 = 0$ and selection probabilities $q_1 = 1 - \delta$, $q_2 = \delta$ respectively.



- Therefore capacity is given by

$$C = \sum_{i=1}^{2} q_i C_i = 1 - \delta \text{ bits/use}$$

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