Channel Capacity and Channel Models

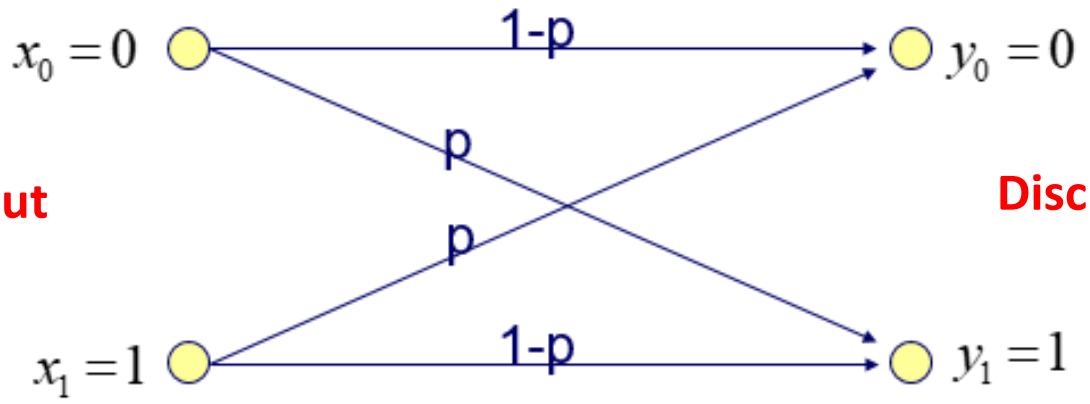
Lecture 13

- In this module, we address and try to find an answer to the questions

  Q1: What happens to information when transmitted over a channel?

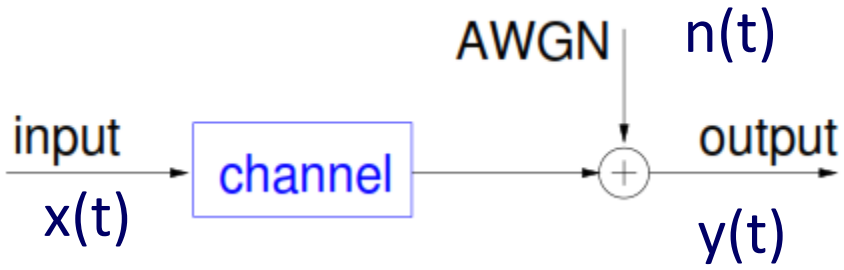
  Q2: What is the maximum rate at which information can be transmitted.

- We will consider two channel models,
  - The Discrete Memory-less channel (in this module)



Discrete Input **Discrete Output**

  - The continuous Gaussian channel (in the next module).



Continuous Input **Continuous Output**

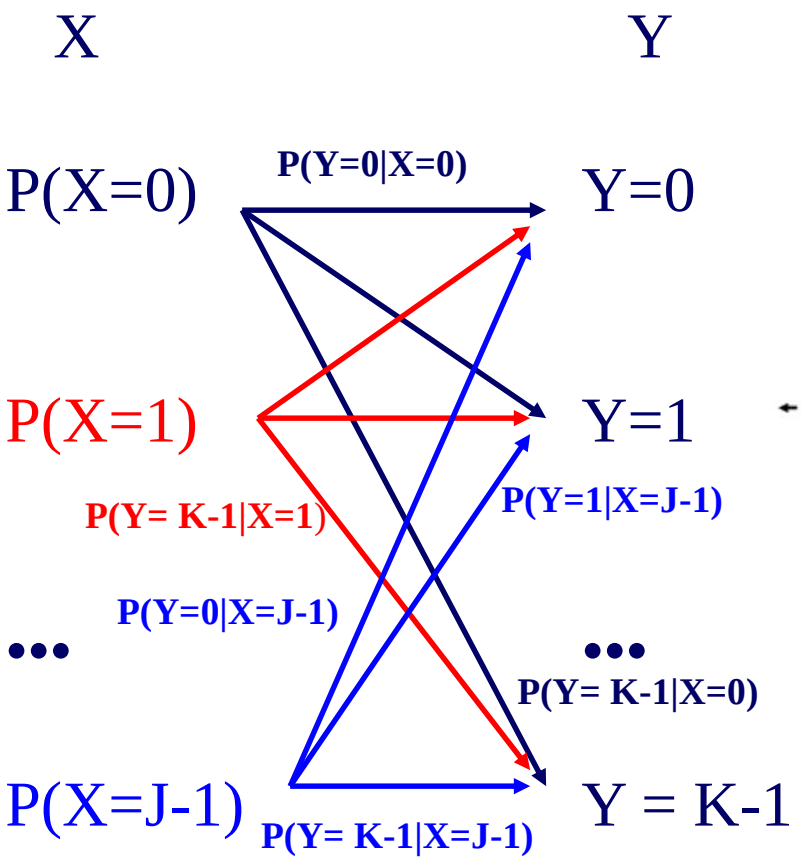
Discrete Memoryless Channel (DMC)

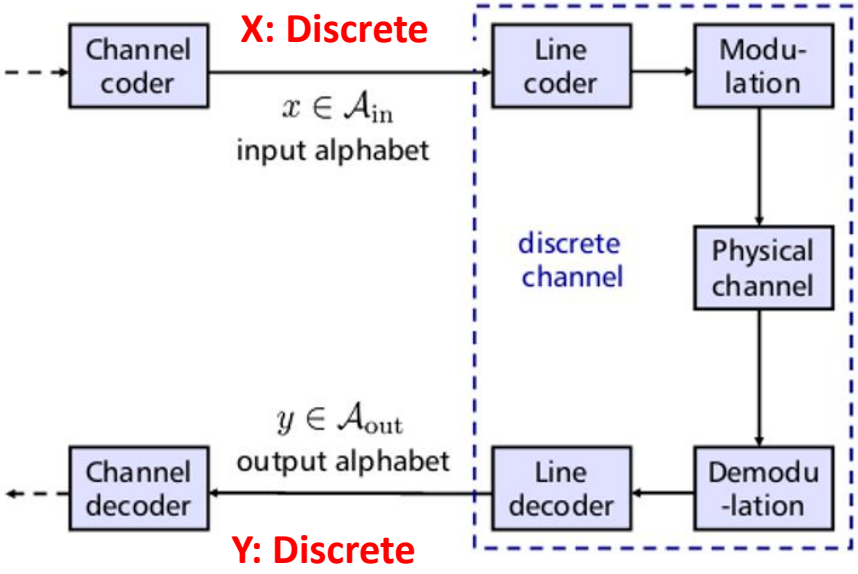
- Definition of DMC
  - Channel with input X (of size **J**) and output Y (of size **K**) which is a noisy version of X. In general, J does not necessarily equal K.
  - Discrete when both X and Y are alphabets of finite sizes.
  - Memoryless when present values of X affect only present values of Y. No dependency on past values of X.
- If $x_1, x_2, \ldots, x_n$ is a sequence of input symbols and $y_1, y_2, \ldots, y_n$ is the corresponding sequence of output symbols, then

$$p(y_1, y_2, \ldots, y_n \mid x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(y_i \mid x_i)$$

Model of a Discrete Channel

Channel is characterized by a set of transition probabilities.





$$P(X = j, Y = k) = P(X = j)P(Y = k \mid X = j)$$

$$P(Y = j) = \sum_{X=0}^{X=J-1} P(X = j)P(Y = k \mid X = j)$$

$$= \sum_{X=0}^{X=J-1} P(X = j, Y = k)$$

STUDENTS-HUB.com

Uploaded By: Mohammad Awawdeh

Discrete Memoryless Channel

- Given the marginal pdf of X, $P(X = x_j)$, and the channel transition probabilities $P(y_k/x_j)$

  - The joint prob. distribution of X and Y

$$p(x_j, y_k) = p(X = x_j \bigcap Y = y_k) = p(X = x_j)\,p(Y = y_k / X = x_j)$$
$$= p(x_j)\,p(y_k / x_j)$$

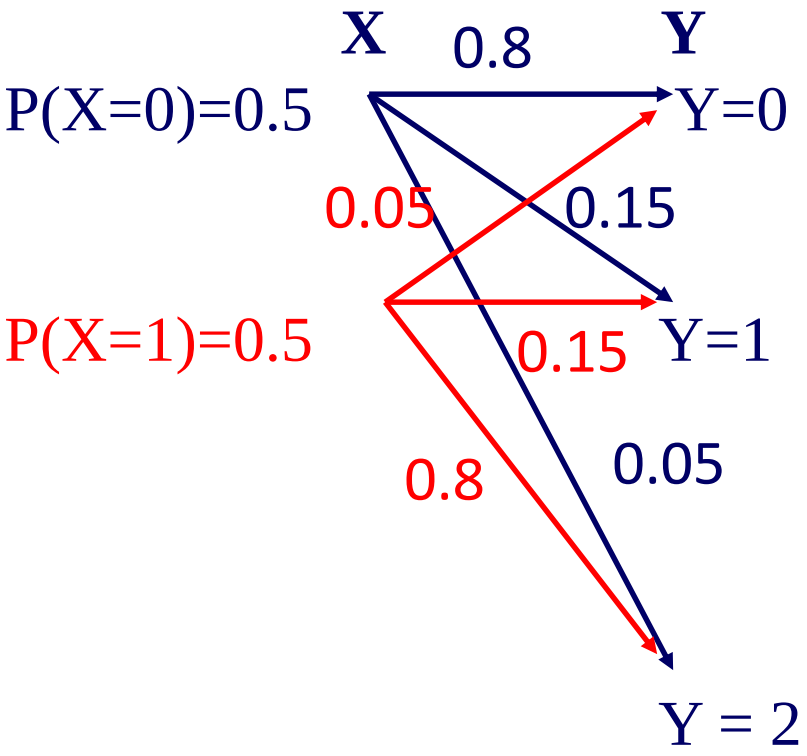
  - The marginal pdf of the output Y,

$$p(y_k) = p(Y = y_k) = \sum_{j=0}^{J-1} p(X = x_j)\,p(Y = y_k / X = x_j)$$
$$= \sum_{j=0}^{J-1} p(x_j)\,p(y_k / x_j), \quad for \quad k = 0,1,..,K-1$$

STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Example: Discrete Memoryless Channel

Consider a DMC with two equally probable input symbols (0, 1) and three output symbols (0, 1, 2). The transition probabilities are as shown in the figure. Find the probability distribution of the channel output.



$$p(x_j, y_k) = p(x_j)\,p(y_k / x_j)$$

$$p(y_k) = \sum_{j=0}^{J-1} p(X = x_j)\,p(Y = y_k / X = x_j)$$

$$P(Y=0) = 0.5*0.8 + 0.5*0.05 = 0.425$$

$$P(Y=1) = 0.5*0.15 + 0.5*0.15 = 0.15$$

$$P(Y=2) = 0.5*0.05 + 0.5*0.8 = 0.425$$

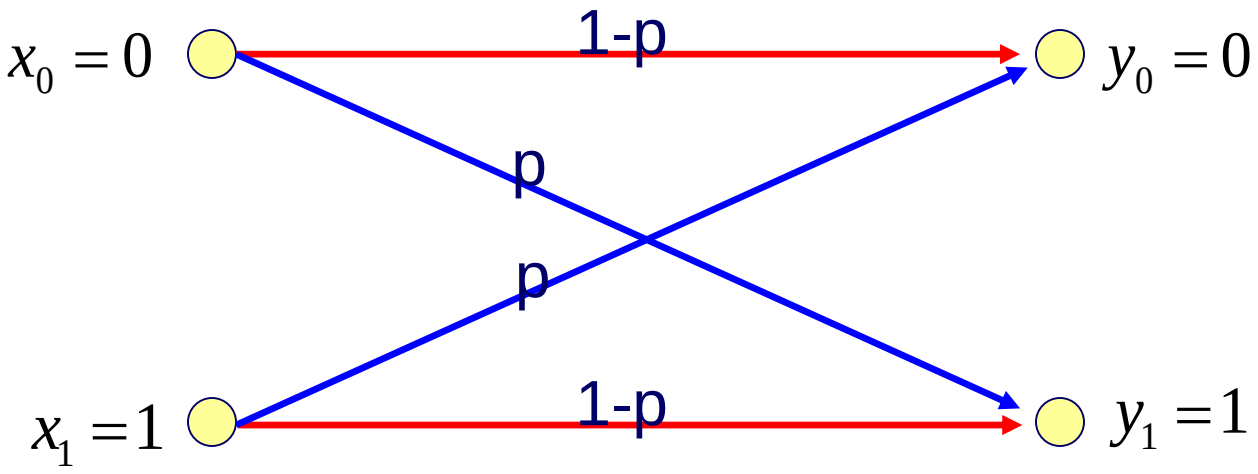
Probabilities sum to 1

STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Binary Symmetric Channel (BSC)

- For the binary symmetric channel, $P(y_0/x_1)=P(y_1/x_0) = p$.

$$P_b^* = Q\left(\sqrt{\frac{\int_0^\tau (s_1(t)-s_2(t))^2 dt}{2N_0}}\right) = p$$



$$P(\mathrm{Y}=0) = P(\mathrm{X}=0)(1-\mathrm{p}) + P(\mathrm{X}=1)\,\mathrm{p}$$

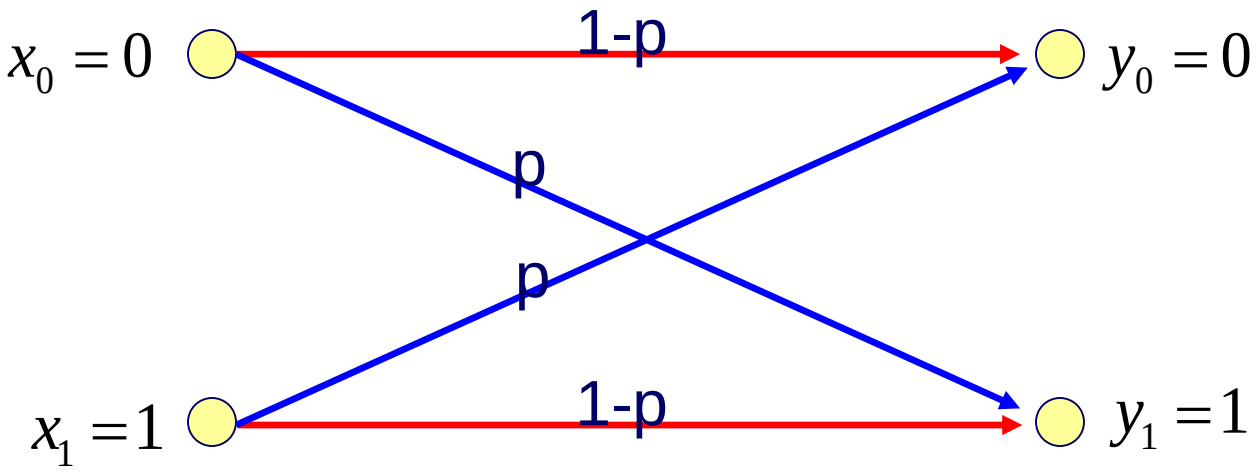
$$P(\mathrm{Y}=1) = P(\mathrm{X}=0)\,\mathrm{p} + P(\mathrm{X}=1)(1-\mathrm{p})$$

STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Example: Binary Symmetric Channel (BSC)

$\boldsymbol{Let\ P(X=0)=0.3\ and\ p=0.1.\ Find}$

- $\boldsymbol{a.\ P(Y=0)}$
- **b. P(Y= 0, 1 / X=0, 0)**
- $\boldsymbol{a.\ P(Y=0)=0.3*0.9+0.7*0.1=0.34}$
- $\boldsymbol{b.\ P(Y=0,1|\ X=0,0)=P(Y=0|\ X=0)P(Y=1|\ X=0)=(1-p)p=0.9*0.1=0.09}$ **(using the memoryless property)**



$$P(\mathrm{Y}=0)=P(\mathrm{X}=0)(1-\mathrm{p})+P(\mathrm{X}=1)\mathrm{p}$$

$$P(\mathrm{Y}=1)=P(\mathrm{X}=0)\mathrm{p}+P(\mathrm{X}=1)(1-\mathrm{p})$$

STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Joint Entropy and Mutual Information

Lecture 14

Review. ***Consider*** **a source S with the following probability distribution:**

symbol	s_1	s_2	⋯	s_M
probability	p_1	p_2	⋯	p_M

- **The entropy H of *S* is defined as:**
- $\mathrm{H}(S) = \sum_{i=1}^{M} -p_i \log_2 p_i$ **(bit/symbol)**
- **The entropy H can be interpreted as**
  - **The average amount of information in the source**
  - **It is a measure of uncertainty in the source**
  - **The minimum number of bits/symbol that is needed to represent the source (the source coding theorem)**

The Joint Entropy

- Let $X = \{x_1, x_2, \ldots, x_J\}$ be the input to a channel and let $\mathrm{Y} = \{y_1, y_2, \ldots, y_K\}$ be the channel output .
- Let $\mathrm{P}(X = x_j, Y = y_k\}$ be the joint pdf of X and Y.
- The joint occurrence of (x_j, y_k) can be considered as a source in a two-dimensional space.
- The joint entropy of *X* and *Y, represents the uncertainty in the joint event (X,Y) and is defined as*:

$$H(X,Y) = E\{\log(1/P(X,Y))\}$$

$$= -\sum_{j=1}^{J-1}\sum_{k=1}^{K-1} P(x_j, y_k)\log_2 P(x_j, y_k)$$

Joint and the Conditional Entropies

An alternative representation of the joint entropy can be found *as*:

$$H(X,Y) = -\sum_{j=1}^{J-1}\sum_{k=1}^{K-1} P(x_j, y_k)\log_2 P(x_j, y_k)$$

$$= -\sum_{x=1}^{J-1}\sum_{k=1}^{K-1} P(x_j)P(y_k / x_j)\log_2 P(x_j)P(y_k / x_j)$$

$$= -\sum_{X} P(x_j)\sum_{Y} P(y_k / x_j)\log_2 P(x_j) - \sum_{X,Y} P(x_j, y_k)\log_2 P(y_k \mid x_j)$$

$$= H(X) + H(Y \mid X)$$

$$H(X,Y) = H(\mathrm{X}) + H(\mathrm{Y} \mid \mathrm{X})$$

$$H(X,Y) = H(Y) + H(X \mid Y)$$

When X and Y are independent (prove)

$$H(X,Y) = H(Y) + H(X)$$

The uncertainty in the joint event (X,Y) is the sum of the uncertainty in X plus the remaining uncertainty in Y after X is known.

The uncertainty in the joint event (X,Y) is the sum of the uncertainties in X and Y when they are independent

Joint and Conditional Entropies: Example

Consider two random variables X and Y with the joint pdf as shown in the table below. Find H(X, Y), H(X), H(Y), H(Y/X) and H(X/Y).

	P(Y = y)	1/2	1/4	1/8	1/8
P(X=x)	Y / X	0	1	2	3
1/4	0	1/8	1/16	1/32	1/32
1/4	1	1/16	1/8	1/32	1/32
1/4	2	1/16	1/16	1/16	1/16
1/4	3	1/4	0	0	0

$$H(X,Y) = -\sum_{j=1}^{J-1}\sum_{k=1}^{K-1} P(x_j, y_k)\log_2 P(x_j, y_k)$$

$$\mathbf{H}(S) = \sum_{i=1}^{M} -p_i \log_2 p_i$$

$$\mathbf{H(X,Y)} = 2 * \left(\frac{1}{8}\right) log_2(8) + 6 * (\frac{1}{16}) log_2(16) + (1/4)\, log_2(4) + 4*(1/32)$$

$$log_2(32) = \left(\frac{3}{4}\right) + \left(\frac{3}{2}\right) + \left(\frac{1}{2}\right) + \left(\frac{5}{8}\right) = 3.375 bits$$

$$\mathbf{H}(Y) = \left(\frac{1}{2}\right) log_2(2) + \left(\frac{1}{4}\right) log_2(4) + 2 * \left(\frac{1}{8}\right) log_2(8) = 1.75\ bits$$

$$\mathbf{H(X)} = 4 * \left(\frac{1}{4}\right) log_2(4) = 2\ bits$$

$$\mathbf{H(X/Y)} = \mathbf{H(X,Y)} - H(Y) = 3.375 - 1.75 = 1.625$$

$$\mathbf{H(Y/X)} = \mathbf{H(X,Y)} - H(X) = 3.375 - 2 = 1.375$$

Note that

$$\mathbf{H(X,Y) > H(X)}$$

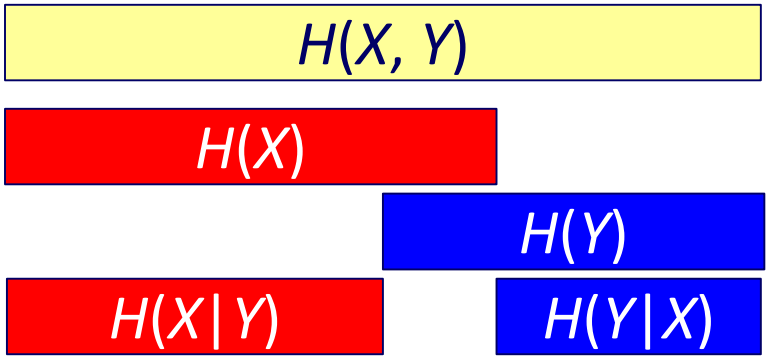
$$\mathbf{H(X,Y) > H(Y)}$$

$$\mathbf{H(X,Y) < H(X) + H(Y)}$$

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Relations among Joint and Conditional Entropies

- Lemma: $H(X,Y) \le H(X) + H(Y)$, (Proof will be given at the end of the video)
- If $P(X,Y) = P(X)P(Y)$, i.e., when X and Y are independent, then
  - $H(X,Y) = H(X) + H(Y)$ (the joint entropy is the sum of the individual entropies).
  - $H(X|Y) = H(X)$ and $H(Y|X) = H(Y)$
-



Observe that:

$H(X|Y) < H(X)$ and

$H(Y|X) < H(Y)$.

Equality holds when X and Y are independent

Mutual Information

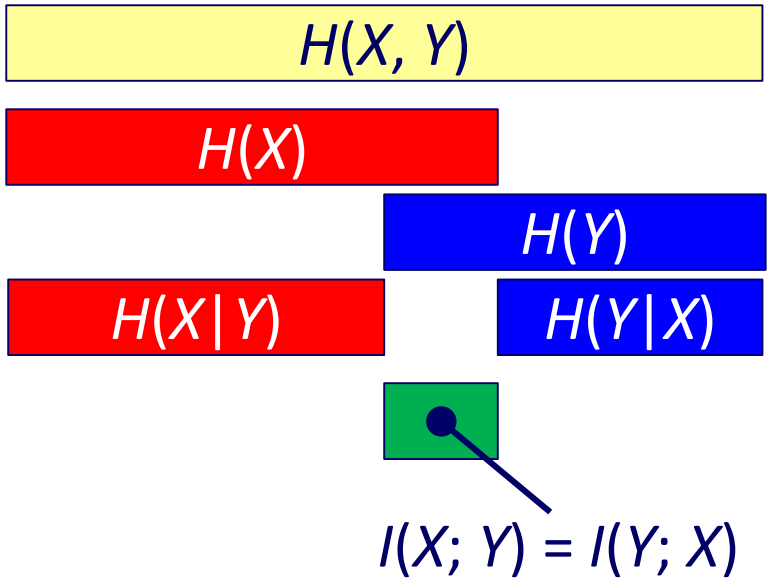
Given two r.v. , X and Y with marginal distributions P(X) and P(Y), **the mutual information** between X and Y is the relative entropy between the joint distribution P(X,Y) and the product distribution P(X)P(Y). It is a measure of the amount of information one r.v. contains about another r.v.

$$I(X,Y) = \sum_{X,Y} P(x,y)\log_2 \frac{P(x,y)}{P(x)P(y)}$$

$$= H(X) - H(X|Y)$$

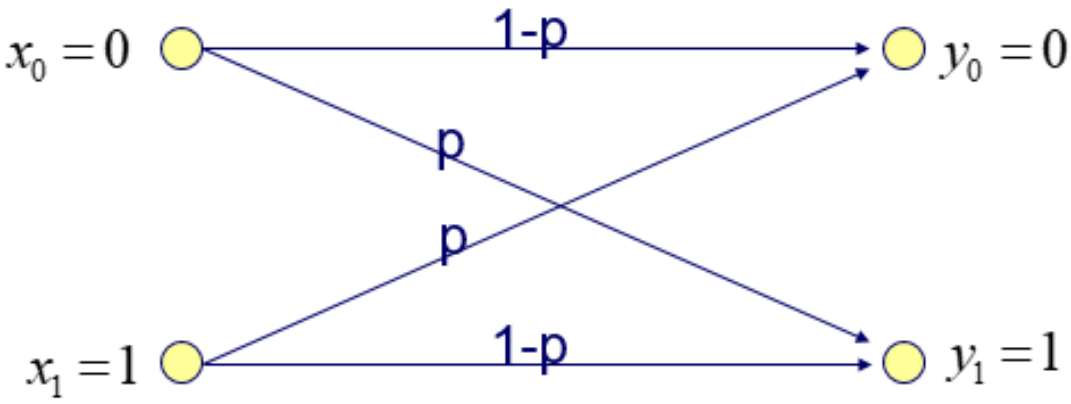
$$= H(Y) - H(Y|X)$$

- H(X): Uncertainty about X
- H(X/Y):Remaining uncertainty about X after Y is being observed
- The difference is the amount of information conveyed by Y (the reduction in uncertainty)



Mutual Information

- Mutual information is the amount of uncertainty about the channel input X resolved on observing the channel output Y.
- For a good channel, one would expect $I(X;Y)$ to be large, i.e., on observing Y, we are able to resolve X with a high degree of reliability.



$$I(X,Y) = H(X) - H(X \mid Y)$$

Joint and Conditional Entropies: Example

Consider two random variables X and Y with the joint pdf as shown in the table below. Find I(X;Y)

	P(Y = y)	1/2	1/4	1/8	1/8
P(X=x)	Y / X	0	1	2	3
1/4	0	1/8	1/16	1/32	1/32
1/4	1	1/16	1/8	1/32	1/32
1/4	2	1/16	1/16	1/16	1/16
1/4	3	1/4	0	0	0

For this example, we obtained earlier that:

$$\mathrm{H}(\mathrm{X},\mathrm{Y}) = 3.375 bits, \mathrm{H}(Y) = 1.75\ bit \quad \mathrm{H}(\mathrm{X}) = 2\ bits$$

Note that $I(X;Y) = H(X) - H(X|Y)$

But, $\mathrm{H}(X|Y) = H(X,Y) - H(Y)$, then

$I(X;Y) = H(X) + H(Y) - H(X,Y)$,

$I(X;Y) = 2 + 1.75 - 3.375 = 0.375$

Entropy: A Proof That H(X) ≥ H(X|Y)

$$H(X) = E\{\log(1/P(X))\} = \sum_{X,Y} P(x,y)\log_2(1/P(x))$$

$$H(X|Y) = E\{\log(1/P(X|Y)) = \sum_{X,Y} P(x,y)\log_2(1/P(x|y))$$

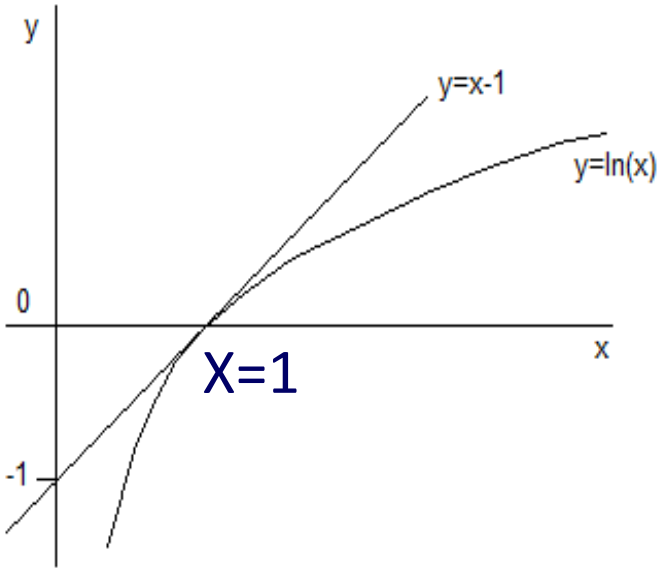
$$H(X) - H(X|Y) =$$

$$= -\sum_{x,y} P(x,y)\log\frac{P(x)}{P(x|y)}$$

$$\geq -\sum_{x,y} P(x,y)\left(\frac{P(x)}{P(x|y)} - 1\right)$$

$$\geq -\sum_{x,y} P(x,y)\left(\frac{P(x)P(y)}{P(x,y)} - 1\right)$$

$$\geq -\sum_{x,y} P(x)P(y) + \sum_{x,y} P(x,y) = 0$$



Uploaded By: Mohammad Awawdeh

Mutual Information

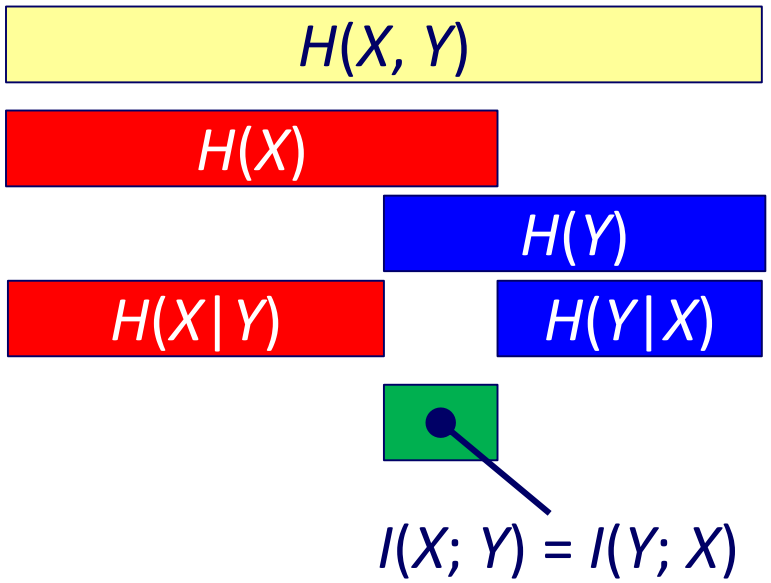
Lecture 15

Given two r.v. , X and Y with marginal distributions P(X) and P(Y), **the mutual information** between X and Y is the relative entropy between the joint distribution P(X,Y) and the product distribution P(X)P(Y). It is a measure of the amount of information one r.v. conveys about another r.v.

$$I(X,Y) = \sum_{X,Y} P(x,y)\log_2 \frac{P(x,y)}{P(x)P(y)}$$

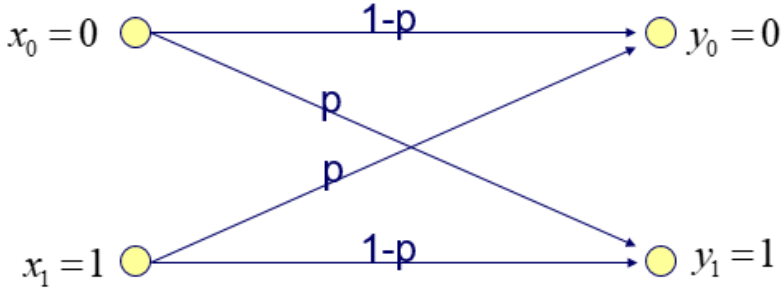
$$= H(X) - H(X \mid Y)$$

$$= H(Y) - H(Y \mid X)$$



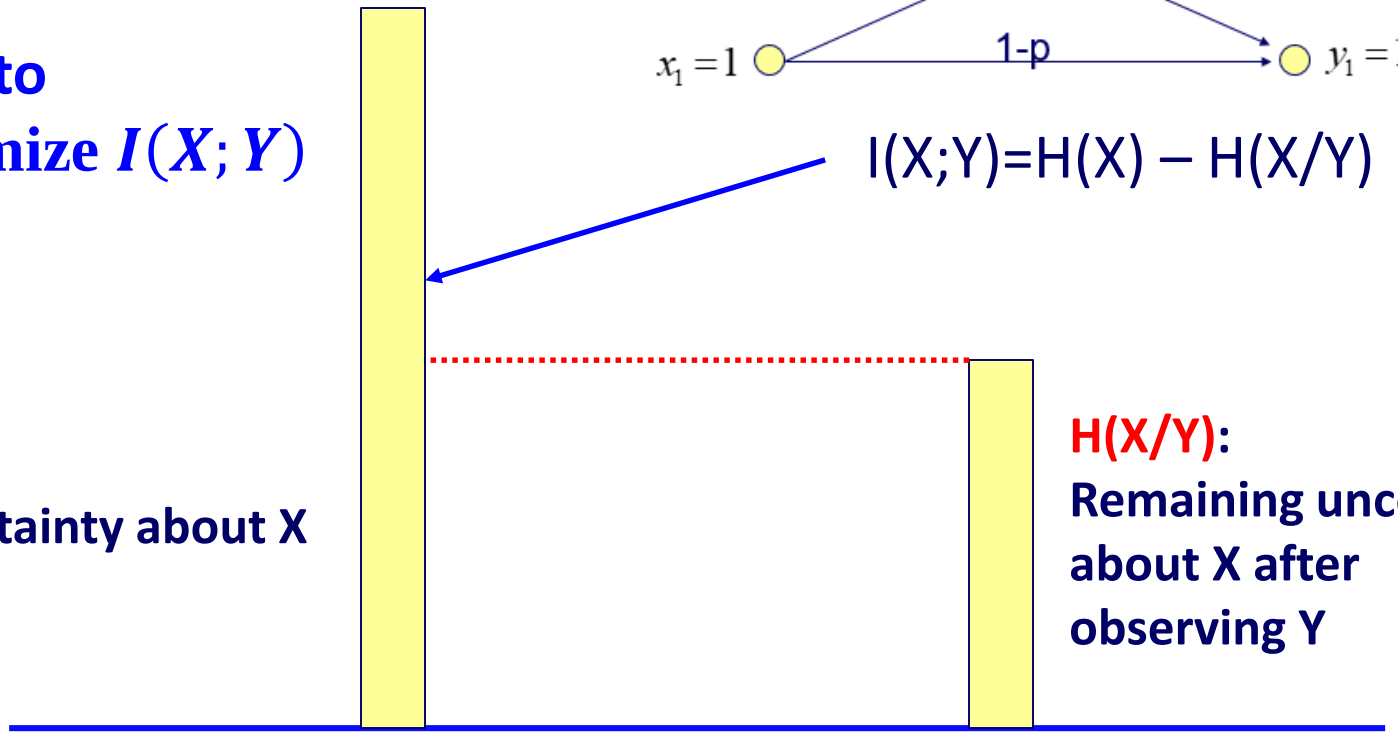
Mutual Information

- Mutual information is the amount of uncertainty about the channel input X resolved upon observing the channel output Y.
- For a reliable channel, one would like to maximize $I(X;Y)$ on observing Y.



Need to maximize $I(X;Y)$

I(X;Y)=H(X) – H(X/Y)



H(X):
uncertainty about X

H(X/Y):
Remaining uncertainty about X after observing Y

Channel Capacity

- For a DMC with input X, output Y, the mutual information between the channel input X and its output Y is

$$I(X;Y)=\sum_{j=0}^{J-1}\sum_{k=0}^{K-1}p(x_j,y_k)\log_2[\frac{p(y_k\mid x_j)}{p(y_k)}]=\sum_{j=0}^{J-1}\sum_{k=0}^{K-1}p(x_j)\,\mathrm{p}(y_k\mid x_j)\log_2[\frac{p(y_k\mid x_j)}{p(y_k)}]$$

Note that: I(X;Y) depends on

a. Input probability distribution $p(x_j)$

b. Transition probabilities $p(y_k|x_j)$. These probabilities depend on the amount of noise present in the channel (usually, not under the control of the user)

- Hence, to maximize the mutual information, one would carry the maximization over the input probability distribution.
- The channel capacity is defined as

$$C = max_{p(X)} I(X;Y) \text{ bits/symbol}$$

- Maximization over all input probability distributions.
- C: Channel Capacity which denotes the maximum rate at which information can be transmitted reliably over the channel.

STUDENTSHUB.com Uploaded By: Mohammad Awawdeh

Example: Capacity of the binary symmetric channel

- Find the capacity of the BSC with cross-over probability p
- To find the capacity, assume P(1)=u, P(0)=1-u. Need to find u that maximizes I(X; Y). **The capacity is the maximum of I(X; Y).**

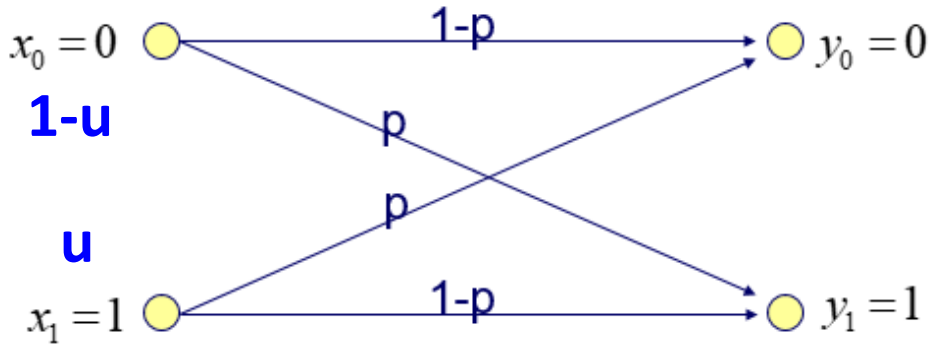
$$I(X;Y) = \sum_{j=0}^{1}\sum_{k=0}^{1} p(x_j)P(y_k \mid x_j)\log_2[\frac{p(y_k \mid x_j)}{p(y_k)}$$

$$I(X;Y) = (1-u)(1-p)\log\frac{1-p}{(1-p)(1-u)+pu}$$

$$+(1-u)p\log\frac{p}{p(1-u)+(1-p)u}$$

$$+up\log\frac{p}{(1-p)(1-u)+pu}$$

$$+(u)(1-p)\log\frac{1-p}{p(1-u)+(1-p)u}$$



$$P(Y=0) = (1-u)(1-p)+up$$

$$P(Y=1) = (1-u)p+u(1-p)$$

STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Example: Capacity of the binary symmetric channel

- To obtain the Channel Capacity (this is very tedious)
  - differentiate I(X; Y) w.r.t. u.
  - Set the derivative to zero .
  - Solve for u. The maximum is attained when u=1/2, i.e., for equally probable input symbols.
  - Substitute u=1/2 into I(X;Y). The result is the channel capacity

$$C = \max\ I(X;Y) = I(X;Y)\big|_{p(x_0)=0.5}$$

$$\therefore C = 1 + p\log_2 p + (1-p)\log_2(1-p) = 1 - h(p)$$

where h(.) is the binary entropy function introduced earlier.

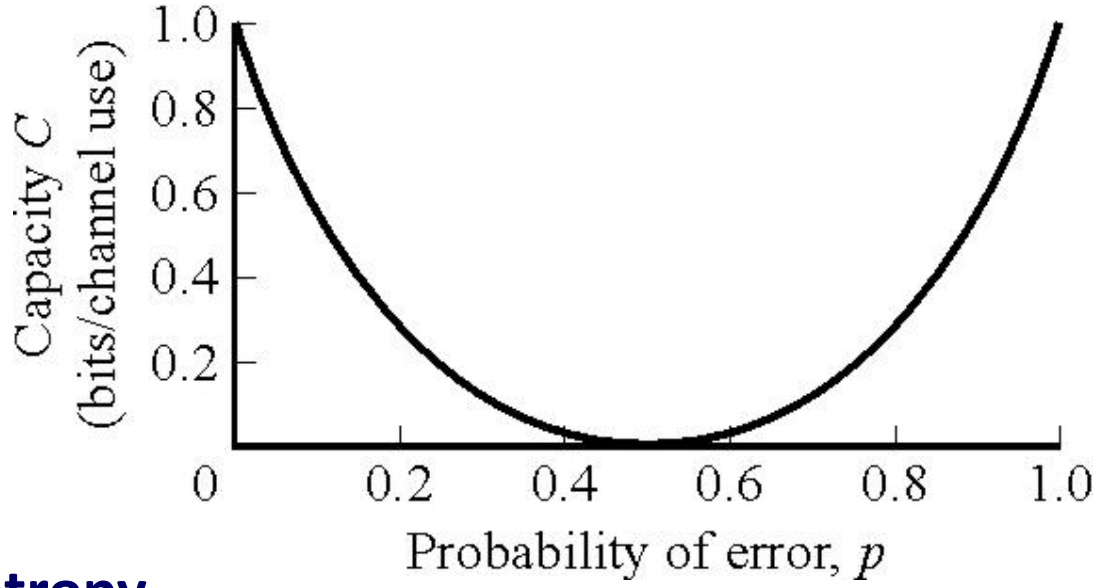
$$\mathrm{h(x)} = -x\log_2 x - (1-x)\log_2(1-x)$$

Example: Capacity of the binary symmetric channel

$$C = 1 + p\log_2 p + (1-p)\log_2(1-p)$$

$$C = 1 - h(p)$$

- Channel Capacity



I(X;Y)=H(X) – H(X/Y)

Remarks:

1. **0 ≤ C ≤1**
2. **C ≤ H(X); the source entropy**
3. **C is max when channel makes no errors, p=0 (noiseless channel).**
1. **When *p*=1 bits are inverted but information is perfect if invert them back!**
4. **Channel conveys no information when p=0.5 (a very noisy channel)**

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Example: Capacity of a Noiseless Channel

- The noiseless channel is deterministic. If you know Y, certainly, you know X (no remaining uncertainty about X after observing Y).
- This is a special case of the BSC when p=0.
- For this channel, Y = X, and therefore, H(X/Y)=0

$$I(X,Y) = H(\mathrm{X}) - H(\mathrm{X}\,|\,\mathrm{Y})$$
$$= H(\mathrm{X})$$

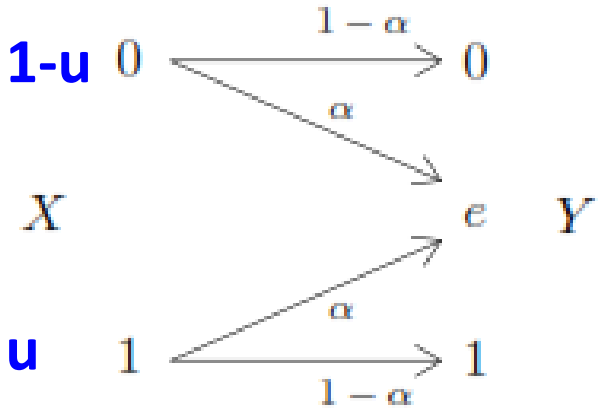
H(X) is maximized when P(X=1)=P(X=0)=1/2. This implies that the capacity of the channel is

C = 1 bit/transmission

Exercise: Capacity of the Erasure channel

- The input consists of the symbols (0,1) and the output consists of the symbols 0, 1, and an erased condition (e)

$$I(X,Y) = H(Y) - H(Y \mid X)$$
$$= H(Y) - h(\alpha)$$
$$H(Y) = (1-\alpha)\mathrm{H}(X) + h(\alpha)$$



By substitution, we get

$$I(X,Y) = (1-\alpha)H(X)$$

$$\mathrm{h(x)} = -x\log_2 x - (1-x)\log_2(1-x)$$

This is maximized when P(X=0)=P(X=1)=1/2. In which case, H(X) =1, when the symbols have equal probabilities. The capacity becomes

$$C = (1-\alpha) \text{ bit/transmission}$$

Shannon's Second Theorem (noisy coding theorem)

Consider a DMS with alphabet X and entropy H(X) that produces symbols at a rate of one symbol every T_s; $i, e, rate\ R_s$ symbols/sec . The output is transmitted over a DMC that has a capacity C bits/ transmission and can be used once every T_c , $i.e. rate\ R_c$ times per sec. Then,

a. if $H(X)\,\mathrm{R}_s < CR_c$ — there exists a coding scheme capable of achieving an arbitrary low probability of error .

b. if $H(X)\,\mathrm{R}_s > CR_c$ — it is not possible to transmit with arbitrary small error

Source Rate= Rs=1/Ts symbols/sec. Each symbol carries H bits of information. Hence, Source information rate=**HRs** bits of information/ sec.

Channel rate= Rc =1/Tc transmissions/sec. Information carried/transmission is the channel capacity C in bits/transmission. Hence, channel conveys RC bits/sec

One source bit (Ts)	One source bit (Ts)	One source bit (Ts)	One source bit (Ts)

k

One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)

n

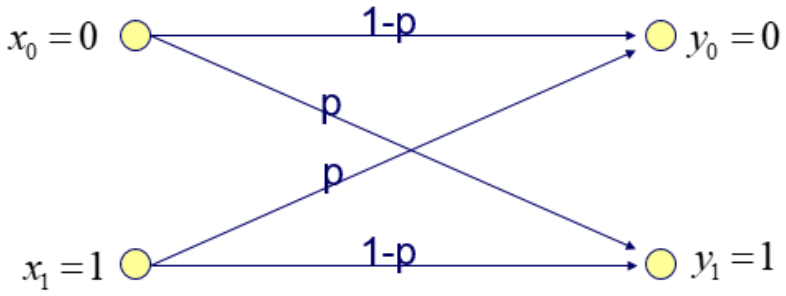
STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Channel Coding Theorem for DMC

Example: Consider a BSC with $p(x_0) = 0.5$. Here, H(X)=1.

Solution: Since H(X)=1, the condition for reliable communication:

$$R_s \leq CR_c \; or \; \frac{R_s}{R_c} \leq C$$



Let us define the code rate as:

$$\boldsymbol{r} = \frac{R_s}{R_c} = \frac{T_c}{T_s}$$

for $\boldsymbol{r} \leq \boldsymbol{C}$, there exists a code (with code rate less than or equal to C) capable of achieving an arbitrary low probability of error.

One source bit (Ts)	One source bit (Ts)	One source bit (Ts)	One source bit (Ts)

k

One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)	One channel bit (Tc)

n

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

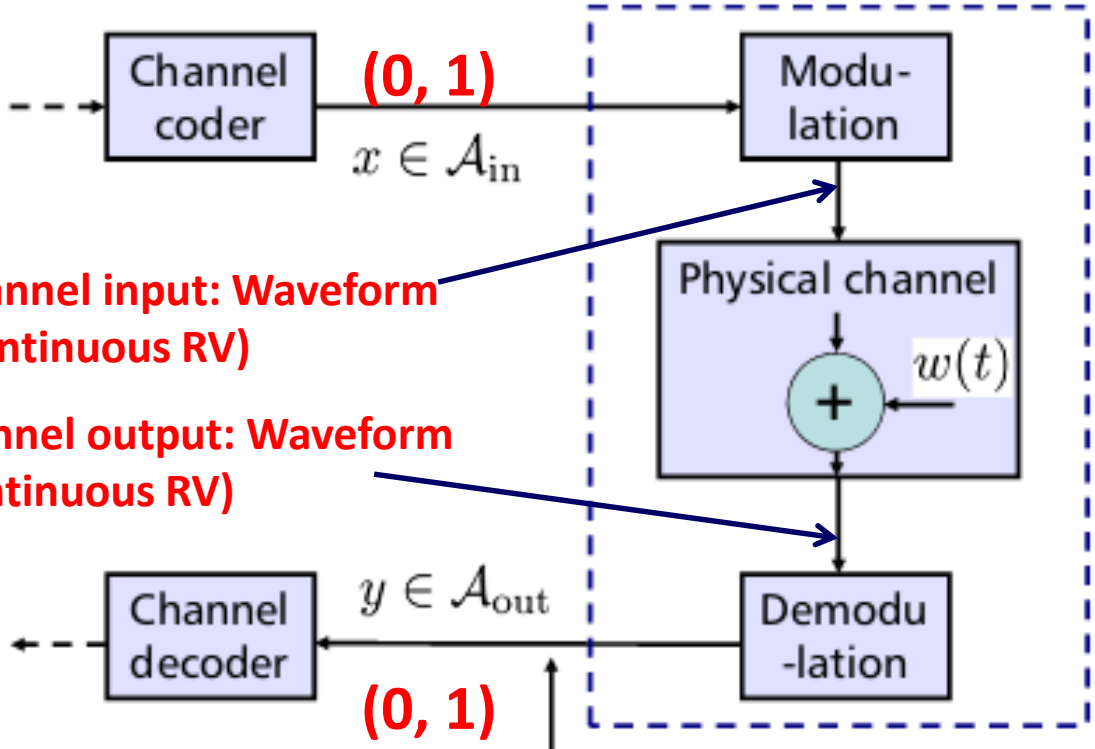
Channel Coding Theorem for DMC

- Suppose r<C, where C is the capacity of a DMC, then for any ε>0, there exists a code of rate r and length n such that the probability of block error decoding < ε when the code is used on the channel.
- The theorem emphasizes the existence of a code, but does not specify the code itself.
- Suppose that we have k independent message bits and we add (n-k) redundant bits derived from the k bits. The result is a codeword of length n. During transmission, there will be errors in the n bits. But the redundant (n-k) bits will serve to reduce the block error probability so we can recover the k message bits with an arbitrarily small probability of error. Later on in the course, we will consider these encoding schemes.
- Theorem will not be proved here.

Capacity of the AWGN Continuous Channel (Modulated waveform channels)

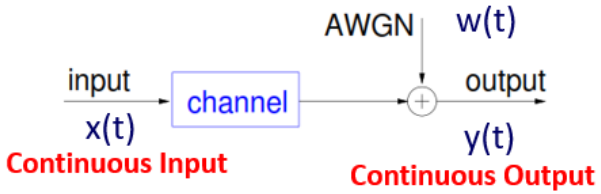
Lecture 16

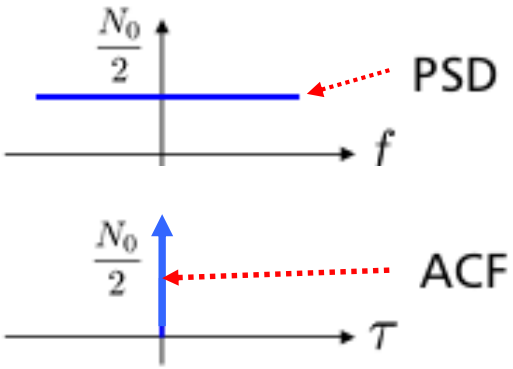
AWGN (Additive White Gaussian Noise) Channel:

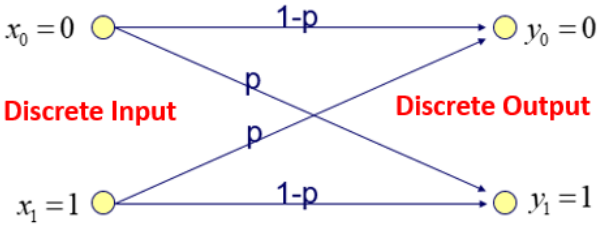


Channel input: Waveform (Continuous RV)

Channel output: Waveform (Continuous RV)







$$\sigma_w^2 = \frac{N_0}{2}$$

Demodulator limits bandwidth. The noise variance at the sampling times computes to $\frac{N_0}{2}$.

Continuous Sources: Differential Entropy

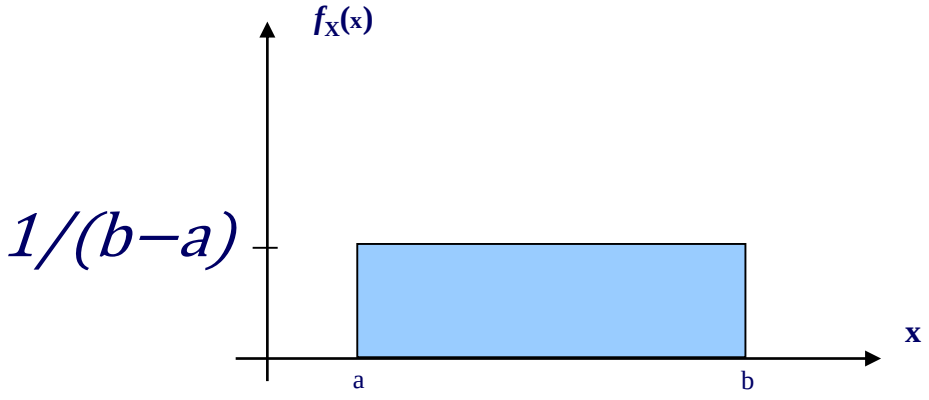
- For a source X with a discrete alphabet, we defined entropy as
- $\boldsymbol{H(X) = -\sum_{x=0}^{M-1} P_x log_2(P_x)}$
- For a source X with a continuous distribution, we define the differential entropy as
- $\boldsymbol{h(X) = -\int_{-\infty}^{\infty} f_X(x) log_2\big(f_X(x)\big) dx}$
- Here, $f_X(x)$ is the probability density function of the random variable X.
- Note that small h is used to denote the differential entropy.

Example: Differential Entropy of a Uniform pdf

- Consider a source X with a uniform distribution as shown below

$$f_X(x) = \begin{cases} 1/(b-a) \ , & a \le x \le b \\ 0 \ , & otherwise \end{cases}$$

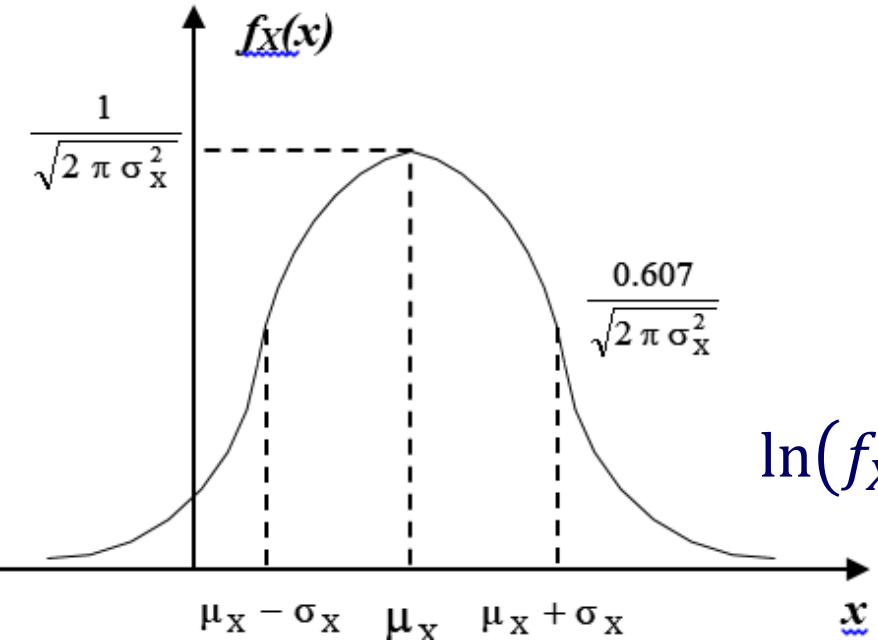
- Let $\Delta = (b-a)$. The differential entropy is
- $h(X) = -\int_{-\infty}^{\infty} f_X(x) log_2\big(f_X(x)\big) dx = -\int_{-\infty}^{\infty} (\frac{1}{\Delta}) log_2\left(\frac{1}{\Delta}\right) dx$
- $h(X) = -(\frac{1}{\Delta}) log_2\left(\frac{1}{\Delta}\right) \Delta = log_2(\Delta)$
- When $\Delta < 1$, $h(X)$ is negative. Hence, unlike entropy, differential entropy can be negative.



Differential Entropy of the Gaussian Source

- Consider a source X with a Gaussian distribution as shown below

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{\frac{-(x-\mu_X)^2}{2\sigma_X^2}}$$



Note that:

$$\ln(ab) = \ln(a) + \ln(b)$$
$$\ln(a/b) = \ln(a) - \ln(b)$$
$$\ln(1/b) = -\ln(b)$$
$$log_2(u) = \ln(u)/\ln(2)$$

Also, $log_2(e) = \ln(e)/\ln(2)$

Hence $\frac{1}{\ln(2)} = log_2(e)$

Therefore, $log_2(u) = log_2(e)\ln(u)$

$$\ln(f_X(x)) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}((x-\mu)^2$$

$$log_2(f_X(x)) = -log_2(e)(\frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}((x-\mu)^2)$$

Differential Entropy of the Gaussian Source

- The differential entropy is
- $h(X) = -\int_{-\infty}^{\infty} f_X(x) log_2\big(f_X(x)\big)dx$
- $log_2\big(f_X(x)\big) = -log_2 e\{\frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}((x-\mu)^2\}$
- $h(X) = \int_{-\infty}^{\infty} f_X(x) log_2 e\{\frac{1}{2}\ln(2\pi\sigma^2)\} +$
- $\int_{-\infty}^{\infty} f_X(x) log_2 e\{\frac{1}{2\sigma^2}((x-\mu)^2\}$
- $h(X) = log_2 e\{\frac{1}{2}\ln(2\pi\sigma^2)\}\int_{-\infty}^{\infty} f_X(x)dx;\ \ \int_{-\infty}^{\infty} f_X(x)dx = 1$
- $+log_2 e\frac{1}{2\sigma^2}\int_{-\infty}^{\infty} f_X(x)((x-\mu)^2\}dx;\ \ \int_{-\infty}^{\infty} f_X(x)((x-\mu)^2\}dx=\sigma^2$
- $h(X) = \frac{1}{2} log_2 e\{\ln(2\pi\sigma^2)\} + \frac{1}{2} log_2 e$
- $h(X) = \frac{1}{2} log_2(2\pi\sigma^2) + \frac{1}{2} log_2 e \Longrightarrow \mathbf{h(X) = \frac{1}{2} log_2\big(2\pi e\sigma^2\big)}$
- $\mathbf{h(X)}$ depends only on the variance. As σ^2 increases, $\mathbf{h(X)}$ increases

STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Definitions: Entropy and Conditional Entropy

Integration replaces summation for the case of continuous distributions

Differential entropies:

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 \left(f_X(x)\right) \mathrm{d}x \qquad h(Y) = -\int_{-\infty}^{\infty} f_X(y) \log_2 \left(f_Y(y)\right) \mathrm{d}y$$

$$h(X|Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log_2 \left(\frac{1}{f_X(x|y)}\right) \mathrm{d}x \; \mathrm{d}y = E\left\{\log_2 \left(\frac{1}{f_X(x|y)}\right)\right\}$$

$$h(Y|X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log_2 \left(\frac{1}{f_X(y|x)}\right) \mathrm{d}x \; \mathrm{d}y = E\left\{\log_2 \left(\frac{1}{f_X(y|x)}\right)\right\}$$

Mutual information:

(i) $I(X;Y) = I(Y;X)$

(ii) $I(X;Y) \geq 0$

(iii) $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$

Theorem: Maximum Differential Entropy for Specified Variance

Optimization Problem:

Find p.d.f. for which $h(x)$ is maximum, subject to two constraints

$$i)\ \int_{-\infty}^{\infty} f_X(x)dx = 1 \quad , \quad ii)\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx = \sigma^2 = const$$

where μ is the mean, and σ^2 is the variance (measure of average power)

Solution: Based on calculus of variation & use of Lagrange multiplier

$$I = \int_{-\infty}^{\infty} \left[-f_X(x)\log_2 f_X(x) + \lambda_1 f_X(x) + \lambda_2 (x-\mu)^2 f_X(x) \right] dx$$

λ_1 and λ_2 are the Lagrange multipliers. The desired form of $f_X(x)$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

The corresponding maximum entropy

$$h(x) = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

Remark: For a discrete rv, a uniform distribution results in maximum entropy. For a continuous rv, the Gaussian distribution yields the highest differential entropy

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Capacity of the AWGN Continuous Channel (Modulated waveform channel

Lecture 17

Main Results from the previous video

1. For a source X with a continuous distribution, we defined the differential entropy as

$$\mathbf{h(X) = -\int_{-\infty}^{\infty} f_X(x) log_2\big(f_X(x)\big)dx;} \qquad \mathbf{f_X(x)\ is\ the\ pdf\ of\ X.}$$

2. If X is a Gaussian R.V. $X \sim N(\mu, \sigma_x^2)$, then its differential entropy is given by:

$$\boldsymbol{h(X) = \frac{1}{2} log_2\big(2\pi e \sigma_x^2\big)}$$

3. Optimization Problem:

The pdf for which $h(x)$ is maximum, subject to two constraints

$$i) \quad \int_{-\infty}^{\infty} f_X(x)dx = 1 \quad , \quad ii) \quad \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx = \sigma^2 = const$$

where μ is the mean, and σ^2 is the variance (measure of average power)

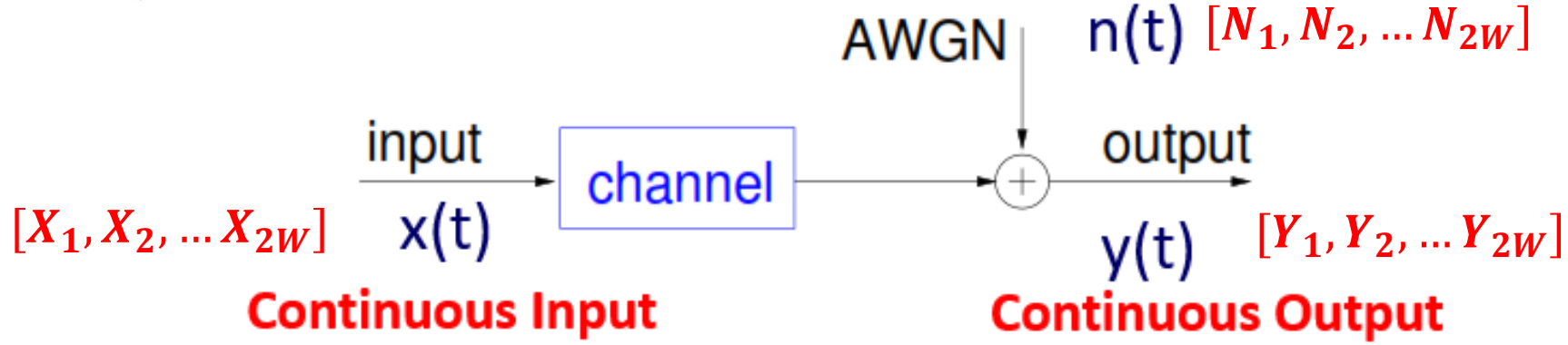
is the Gaussian density function

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

The AWGN Channel Continuous-Input Continuous-Output Channel

- Let us consider the continuous channel with input X and output Y. Additive white Gaussian noise N is added to X during transmission over the channel.
- First, we consider the capacity when X, Y, N are random variables (representing samples from the waveforms x(t), y(t), and n(t).
- The time functions will be treated later
- The noise N is a zero-mean R.V following the Gaussian distribution $N(\mu, \sigma_N^2)$
- The channel input X is a zero-mean random variable with variance σ_X^2
- The objective is to find the capacity of this channel defined as:
- $\boldsymbol{C = max_{f(x)} I(X; Y)}$



STUDENTS-HUB.com
Uploaded By: Mohammad Awawdeh

Mutual Information for a Continuous Channel

The channel input X and Channel output Y are related by:

Y = X + N; X and Y are statistically independent.

The mutual information $\mathrm{I(X;Y)} = \mathrm{h(Y)} - h(Y|X)$

For a given X, $Y = constant(x_0) + N$

But, N$\sim N(0, \sigma_N^2)$, Hence, $\mathrm{f}(Y|X = x_0) \sim N(x_0, \sigma_N^2)$

Earlier, we found that the differential entropy of normal distribution with meanx_0, $and\ variance\ \sigma_N^2$

$$h(Y|X = x_0) = h(N) = \frac{1}{2} log(2\pi e\sigma_N^2) \Longrightarrow h(Y|X) = E_x\{\frac{1}{2} log(2\pi e\sigma_N^2)\}$$

Therefore, $h(Y|X) = \frac{1}{2} log(2\pi e\sigma_N^2)$

Differential entropy in Y/X = Differential entropy in $N(0, \sigma_N^2)$

$$I(X;Y) = h(Y) - h(N) = h(Y) - \frac{1}{2} log(2\pi e\sigma_N^2)$$

STUDENTS-HUB.com

Uploaded By: Mohammad Awawdeh

Capacity of the Continuous Channel

Recall, Y = X + N;

Also, I(X; Y) = h(Y) – h(Y|X)

$$\mathrm{I(X;Y)} = \mathrm{h(Y)} - h(N) = h(Y) - \frac{1}{2}\log(2\pi e\sigma_N^2)$$

- We observe that in order to maximize I(X; Y) we need h(Y) to be maximum, and this happens when Y is Gaussian.
- **Y = X + N**; is Gaussian when both X and N are Gaussian. But N is already Gaussian, so X should be Gaussian.
- Hence, let $\mathrm{X} \sim N(0, \sigma_x^2) = N(0, P_x)$
- Since Y = X + N, then E(Y) = E(X) + E(N),
- Var(Y) = Var(X) + Var(N), $\sigma_Y^2 = \sigma_X^2 + \sigma_N^2 \Longrightarrow h(Y) = \frac{1}{2}\log(2\pi e(\sigma_X^2 + \sigma_N^2)$

C = h(Y) - h(Y|X)= $\frac{1}{2}\log(2\pi e(\sigma_X^2 + \sigma_N^2) - \frac{1}{2}\log(2\pi e\sigma_N^2)$

$$\boldsymbol{C = \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)}$$

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Capacity of the Continuous Channel

- $C = \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)$ in bits/channel use.
- How to find the capacity in bits/sec?
- Now let the bandwidth of the channel be W Hz. By virtue of the sampling theorem, a band-limited signal is completely characterized by its Nyquist sampling rate of 2W samples/sec.
- Hence assume that we transmit 2W samples , X_i, of x(t), to each sample a noise N_i is added to produce the output sample $Y_i = X_i + N_i$;
- $[Y_1, Y_2, \ldots Y_{2W}] = [X_1, X_2, \ldots X_{2W}] + [N_1, N_2, \ldots N_{2W}]$; Samples taken over one second
- The maximum mutual information between each sample X_i and Y_i is
- $C_i = \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)$ bits/symbol.

Therefore, the maximum mutual information between the vectors

$[X_1, X_2, \ldots X_{2W}]$ and $[Y_1, Y_2, \ldots Y_{2W}]$

$C = \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)$ bits/symbol X 2W symbols/sec

$C = W\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)$ bits/se; This is known as the Shannon-Hartley Law. This is one of the fundamental results in modern communication theory.

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Capacity of the Continuous Channel

- $\boldsymbol{C = W\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)}$ bits/sec
- Channel capacity increases as the channel bandwidth increases.
- Channel capacity increases as the signal power to noise ratio increases.
- For a channel with bandwidth W and noise power spectral density $\boldsymbol{N_0/2}$, the noise power $\boldsymbol{\sigma_N^2 = WN_0}$. This can be substituted into the channel capacity formula to obtain:
- $\boldsymbol{C = W\log\left(1 + \frac{\sigma_X^2}{WN_0}\right)}$ bits/sec
- The theorem relates the channel bandwidth, the signal power and the noise power in one formula.
- Sets a theoretical limit on the amount of data that can transmitted reliably over a noisy channel with a given bandwidth.

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Capacity of the Continuous Channel

- $\boldsymbol{C = W\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)}$ bits/sec
- $\boldsymbol{C = W\log\left(1 + \frac{\sigma_X^2}{WN_0}\right)}$ bits/sec
- The theorem sets a limit on the amount of data that can transmitted reliably over a noisy channel with a given bandwidth.
- By employing sophisticated channel encoding algorithms, it is possible to transmit data with arbitrarily small probability of error as long as the data rate is below the channel capacity.
- If data is transmitted at a rate greater than C, then regardless of any encoding scheme employed, there will be a definite probability of error.

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Example 1: An Extremely Noisy Channel

- Consider an extremely noisy channel in which the value of the signal-to-noise ratio is almost zero. In other words, the noise is so strong that the signal is faint. For this channel the capacity C is calculated as
- $\boldsymbol{C = W\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)}$ bits/sec
- $\boldsymbol{C = W\log(1 + 0)}$ = 0 bits/sec
- This means that the capacity of this channel is zero regardless of the bandwidth. In other words, we cannot receive any useful data through this noisy channel.

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh

Example 2: Capacity of the Telephone Line

- Here, we calculate the theoretical highest bit rate of a regular telephone line. A telephone line normally has a bandwidth of 3000 Hz. The signal-to-noise ratio is usually 35dB (3162)
- For this channel, the capacity is calculated as
- $\boldsymbol{C = W\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right)}$ bits/sec
- $\boldsymbol{C = 3000\log\left(1 + 3162\right)}$ = 34,860 bits/sec
- This means that the highest bit rate for a telephone line is 34,860 bps.
- If we want to send data faster than this rate, we can either increase the bandwidth of the line or improve the signal-to-noise ratio.

STUDENTS-HUB.com Uploaded By: Mohammad Awawdeh