

Information Theory and Source Coding

Introduction : Information theory provides a quantitative measure of the information contained in message signal and allows us to determine the capacity of a communication system to transfer this information from source to destination. In this section we briefly explore some basic ideas involved in information theory and source coding.

Introduction to Information Theory

- Information theory is a process that focuses on the task of quantifying information.
- The quantification of information is achieved by identifying viable methods of compressing and communicating data without causing and degradation in the integrity of the data.
- Information theory can be utilized in a number of different fields, including quantum computing, data analysis and cryptography.

Introduction to Information Theory

- The origin of modern informational theory is usually attributed to Claude E. Shannon.
- His work A Mathematical Theory of Communication, first published in 1948, lays the foundation for the quantification and compression of data into viable units that may be stored for easy retrieval later.
- His basic approach provided the tools necessary to enhance the efficiency of early mainframe computer systems, and translated easily into the advent of desktop computers during the decade of the 1970s.

Introduction to Information Theory

- As a branch of both electrical engineering and applied mathematics, information theory seeks to uncover the most efficient methods of conveying data, within the limits inherent in the data proper.
- The idea is to ensure that the mass transit of data does not in any way decrease the quality, even if the data is compressed in some manner.

Introduction to Information Theory

- Ideally, the data can be restored to its original form upon reaching its destination.
- In some cases, however, the goal is to allow data in one form to be converted for mass transmission, received at the point of termination, and easily converted into a format other than the original without losing any of the transmitted information.

What is Information?

- What is meant by the “information” contained in an event?
- If we are formally to define a quantitative measure of information contained in an event, this measure should have some intuitive properties such as:
 1. Information contained in events ought to be defined in terms of some measure of the uncertainty of the events.
 2. Less certain events ought to contain more information than more certain events.
 3. The information of unrelated / independent events taken as a single event should equal the sum of the information of the unrelated events.
- A natural measure of the uncertainty of an event is the probability of A denoted $P(A)$.

Axioms for Information theory

Note that $I(x_i)$ satisfies the following properties;

- $I(x_i) = 0$ for $P(x_i) = 1$
- $I(x_i) \geq 0$
- $I(x_i) > I(x_j)$ if $P(x_i) < P(x_j)$
- $I(x_i, x_j) = I(x_i) + I(x_j)$ if x_i and x_j are independent. (This is based on laws of logarithms.)

Average Information or Entropy

- In a practical communication system, we usually transmit long sequences of symbols from an information source.
- Thus, we are more interested in the average information that a source produces than the information content of a single symbol.
- The mean value of $l(x_i)$ over the alphabet of source X with n different symbols is given by

Units of Measurement

- The unit of $I(x)$ is the bit (binary unit) if $b = 2$, Hartley (or alternatively decit) if $b = 10$, and nat (*natural unit*) if $b = e$ (i.e. the exponential number). We will use $b = 2$.
- Here the unit bit (abbreviated “b”) is a measure of information content and is not to be confused with the term ***bit*** meaning “binary digit.”
- The conversion of these units to other units can be achieved by the following relationships.

$$\log_2 A = \frac{\log_e A}{\log_e 2} = \frac{\log_{10} A}{\log_{10} 2}$$

Remark: $\log_e A$ is also written $\ln A$.

Average Information or Entropy

- In a practical communication system, we usually transmit long sequences of symbols from an information source.
- Thus, we are more interested in the average information that a source produces than the information content of a single symbol.
- The mean value of $I(x_i)$ over the alphabet of source X with n different symbols is given by

$$H(X) = E[I(x_i)] = \sum_{i=1}^m P(x_i)I(x_i)$$

$$H(X) = - \sum_{i=1}^m P(x_i) \log_2(P(x_i)) \text{ (b/symbol)}$$

How to measure information?

A good measure on information content should base on the degree of its uncertainty. From probabilistic viewpoint, if an event is less likely to happen, it should carry more information when it occurs, because it is more uncertain that the event would happen.

In addition, it is reasonable to have additivity for information measure, i.e., the degree-of-uncertainty of a joint event should equal the sum of the degree-of-uncertainty of each individual event. Moreover, a small change in event probability should only yield a small variation in event uncertainty.

How to measure information?

For example, two events respectively with probabilities 0.20001 and 0.19999 should reasonably possess comparable information content.

As it turns out, the only measure satisfying these demands is the self-information denoted as the logarithm of the reciprocal of the event probability .

It is then legitimate to adopt the entropy—the expected value of the self-information—as a measure of information.

How to measure information?

In the case of data transmission over noisy channel, the concern is different from that for data storage (or error-free transmission).

The sender wishes to transmit to the receiver a sequence of pre-defined information symbols under an acceptable symbol error rate. Code redundancies are therefore added to combat the noise.

For example, one may employ the three-times repetition code (i.e., transmit 111 for information symbol 1 and send 000 for information symbol 0), and apply the majority law at the receiver end so that one-bit error can be recovered.

Self-information

Let E be an event with probability $\Pr(E)$, and let $I(E)$ represent the amount of information you gain when you learn that E has occurred (or equivalently, the amount of uncertainty you lose after learning that E has happened). Then a natural question to ask is “what properties should $I(E)$ have?”

The answer to the question may vary person by person. Here are some common properties that $I(E)$, which is called the self-information, is reasonably expected to have.

Measure of Information

1) Information sources:

An information source is an object that produces an event, the outcome of which is selected at random according to a probability distribution.

A practical source in a communication system is a device that produces messages, and it can be either analog or discrete (we deal mainly with the discrete sources, since analog sources can be transformed to discrete sources)

A discrete information source is a source that has only a finite set of symbols as possible outputs. The set of source symbols is called the *source alphabet*, and the elements of the set are called *symbols* or *letters*.

Measure of Information

1) Information sources:

An information source is an object that produces an event. the outcome of which is selected at random according to a probability distribution. A practical source in a communication system is a device that produces messages. and it can be either analog or discrete.

In this chapter we deal mainly with the discrete sources, since analog sources can be transformed to discrete sources through the DSC of sampling and quantization techniques, described in Chap. 5.

A discrete information source is a source that has only a finite set of symbols as possible outputs. The set of source symbols is called the **source alphabet**, and the elements of the set are called *symbols* or *letters*.

Measure of Information

1) Information sources:

An information source is an object that produces an event. the outcome of which is selected at random according to a probability distribution. A practical source in a communication system is a device that produces messages. and it can be either analog or discrete.

In this chapter we deal mainly with the discrete sources, since analog sources can be transformed to discrete sources through the DSC of sampling and quantization techniques, described in Chap. 5.

A discrete information source is a source that has only a finite set of symbols as possible outputs. The set of source symbols is called the **source alphabet**, and the elements of the set are called *symbols* or *letters*.

Memory

- Information sources can be classified as having memory or being memoryless.
- A source with memory is one for which a current symbol depends on the previous symbols.
- A memoryless source is one for which each symbol produced is independent of the previous symbols.
- A discrete memoryless sources (DMS) can be characterized by the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source.

Information content of a Discrete Memoryless Source

The amount of information contained in an event is closely related to its uncertainty. Messages containing knowledge of high probability of occurrence convey relatively little information. We note that if an event is certain (that is, the event occurs with probability 1), it conveys zero information.

Thus, a mathematical measure of information should be a function of the probability of the outcome and should satisfy the following axioms:

1. Information should be proportional to the uncertainty of an outcome.
2. Information contained in independent outcomes should add.

Information content of a Discrete Memoryless Source

- The amount of information contained in an event is closely related to its uncertainty.
- Messages containing knowledge of high probability of occurrence convey relatively little information.
- We note that if an event is certain (that is, the event occurs with probability of 1), then we can say that it conveys zero *information*.
- Conversely - very unlikely events are “high information” events (e.g. Alarms).

Information Content of a Symbol

- Consider a DMS, denoted by X , with alphabet x_1, x_2, \dots, x_n .
- The information content of a symbol x_i , denoted by $I(x_i)$, is defined by

$$I(x_i) = \log_b \left(\frac{1}{P(x_i)} \right) = -\log_b [P(x_i)]$$

where $P(x_i)$ is the probability of occurrence of symbol x_i .

- (We will discuss what b is shortly.)

Information Content of a Symbol:

Consider a DMS, denoted by X , with alphabet x_1, x_2, \dots, x_n . The information content of a symbol x_i , denoted by $I(x_i)$, is defined by

$$I(x_i) = \log_b\left(\frac{1}{P(x_i)}\right) = -\log_b(P(x_i))$$

where $P(x_i)$ is the probability of occurrence of symbol x_i .

The unit of $I(x)$ is the bit (binary unit) if $b \geq 2$, Hartley or decit if $b = 10$, and nat (natural unit) if $b \geq e$. It is standard to use $b \geq 2$. Here the unit bit (abbreviated "b") is a measure of information content and is not to be confused with the term bit meaning "binary digit." The conversion of these units to other units can be achieved by the following relationships.

Information Content of a Symbol:

Consider a DMS, denoted by X , with alphabet x_1, x_2, \dots, x_n . The information content of a symbol x_i , denoted by $I(x_i)$, is defined by

$$I(x_i) = \log_b\left(\frac{1}{P(x_i)}\right) = -\log_b(P(x_i))$$

where $P(x_i)$ is the probability of occurrence of symbol x_i .

Note that $I(x_i)$ satisfies the following properties;



Entropy

- The quantity $H(X)$ is called the entropy of source X .
- It is a measure of the average information content per random symbol.
- The source entropy $H(X)$ can be considered as the average amount of uncertainty within source X that is resolved by use of the alphabet.
- Note that for if binary source X that generates independent symbols 0 and 1 with equal probability, the source entropy $H(X)$ is
- The source entropy $H(X)$ satisfies the following relation: $0 \leq H(X) \leq \log_2 n$ in (10.9) where n is the size (number of symbols) of the alphabet of source X .
- The lower bound corresponds to no uncertainty, which occurs when one symbol has probability $P(x_i) = 1$ while $P(x_j) = 0$ for $j \neq i$, so X emits the same symbol x_i , all the time.
- The upper bound corresponds to the maximum uncertainty which occurs when $P(x_i) = 1/n$ for all i . that is, when all symbols are equally likely to be emitted by X .

Entropy

- The quantity $H(X)$ is called the *entropy* of source X .
- It is a measure of the average information content per random symbol.
- The source entropy $H(X)$ can be considered as the average amount of uncertainty within source X that is resolved by use of the alphabet.
- Note that for if binary source X that generates independent symbols 0 and 1 with equal probability, the source entropy $H(X)$ is

$$H(X) = -1/2\log_2(1/2) - 1/2\log_2(1/2) \text{ b/symbol}$$

Entropy

- The source entropy $H(X)$ satisfies the following relation:

$$0 \leq H(X) \leq \log_2(m)$$

where m is the size (number of symbols) of the alphabet of source X).

- The lower bound corresponds to no uncertainty, which occurs when one symbol has probability $P(x_i) = 1$ (i.e. X emits the same symbol all the time.
- The upper bound corresponds to the maximum uncertainty which occurs when $P(x_i) = 1/m$ for all i . that is, when all symbols are equally likely to be emitted by X .

Entropy: Example

A DMS X has four symbols x_1, x_2, x_3, x_4 with probabilities $P(x_1) = 0.4, P(x_2) = 0.3, P(x_3) = 0.2, P(x_4) = 0.1$.

- (a) Calculate $H(X)$.
- (b) Find the amount of information contained in the messages $x_1x_2x_1x_3$ and $x_4x_3x_3x_2$.

Entropy: Example part a

$$H(X) = - \sum_{i=1}^4 P(x_i) \log_2[P(x_i)]$$

$$H(X) = -0.4\log_2(0.4) - 0.3\log_2(0.3) - 0.2\log_2(0.2) - 0.1\log_2(0.14)$$

$$H(X) = 0.5288 + 0.5210 + 0.4644 + 0.3322 = 1.85 \text{ b/sec}$$

Entropy: Example part b

- (Remark: from probability, recall independent events)
- $P(x_1x_2x_1x_3) = 0.4 \times 0.30 \times 0.40 \times 0.20 = 0.0096$
- $I(x_1x_2x_1x_3) = -\log_2(0.0096) = 6.70\text{b/symbol}$

Entropy: Example part c

- $P(x_4x_3x_3x_2) = 0.1 \times 0.20 \times 0.20 \times 0.30 = 0.0012$
- $I(x_4x_2x_3x_3) = -\log_2(0.0012) = 9.70\text{b/symbol}$

Discrete memoryless channel (from last lecture)

- For a DMC with “m” inputs and “n” outputs, the input X consists of input symbols x_1, x_2, \dots, x_m .
- The probabilities of these source symbols $P(x_i)$ are assumed to be known.
- The output Y consists of output symbols $\{y_1, y_2, \dots, y_n\}$
- Each possible input-to-output path is indicated along with a conditional probability $P(y_j|x_i)$, where $P(y_j|x_i)$ is the conditional probability of obtaining output y_j given that the input is x_i .
- $P(y_j|x_i)$ is called a *channel transition probability*.

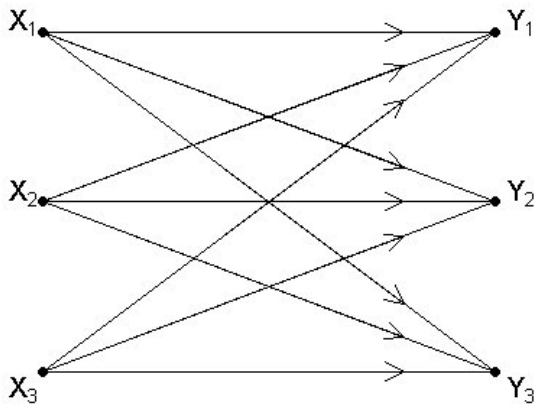
Discrete memoryless channel

- On the next slide, we present a binary DMC, with the channel transition probabilities indicated.
- $P(y_1|x_1) = 0.9$ and $P(y_2|x_1) = 0.1$
- $P(y_1|x_2) = 0.2$ and $P(y_2|x_2) = 0.8$

Discrete Memoryless Channels

- A communication channel is the path or medium through which the symbols flow to the receiver.
- A discrete memoryless channel (DMC) is a statistical model with an input X and an output Y . During each unit of the time, the channel accepts an input symbol from X , and in response it generates an output symbol from Y .
- The channel is “discrete” when the alphabets of X and Y are both finite.
- It is “memoryless” when the current output depends on only the current input and not on any of the previous inputs.

Discrete Memoryless Channels



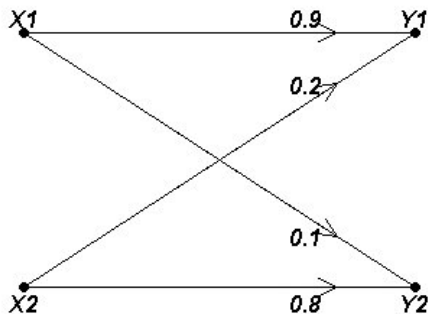
Discrete memoryless channel

- A DMC can have any number of inputs and any number of outputs.
- For a DMC with “m” inputs and “n” outputs, the input X consists of input symbols x_1, x_2, \dots, x_m .
- The probabilities of these source symbols $P(x_i)$ are assumed to be known.
- The output Y consists of output symbols $\{y_1, y_2, \dots, y_n\}$
- Each possible input-to-output path is indicated along with a conditional probability $P(y_i|x_i)$, where $P(y_i|x_i)$ is the conditional probability of obtaining output y_i given that the input is x_i .
- $P(y_i|x_i)$ is called a ***channel transition probability***.

Discrete memoryless channel

- On the next slide, we present a binary DMC, with the channel transition probabilities indicated.
- $P(Y_1|X_1) = 0.9$ and $P(Y_2|X_1) = 0.1$
- $P(Y_1|X_2) = 0.2$ and $P(Y_2|X_2) = 0.8$

Discrete Memoryless Channels



Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, a channel is specified by the matrix of transition probabilities $[P(Y|X)]$, given by

$$[P(Y|X)] = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) & \dots & P(y_n|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & \dots & P(y_n|x_2) \\ \dots & \dots & \dots & \dots \\ P(y_1|x_m) & P(y_2|x_m) & \dots & P(y_n|x_n) \end{bmatrix}$$

The matrix $[P(Y|X)]$ is called the ***channel matrix***.

Channel Matrix

- Since each input to the channel results in some output, each row of the channel matrix must sum to unity (i.e. all rows must add up to 1. This condition is not necessary for columns).
- For the binary DMC presented previously, the channel matrix is

$$[P(Y|X)] = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- (Remark: This is not a binary symmetric channel (relevant to later material))

Channel Matrix

- The input probabilities $P(X)$ are represented by the row matrix

$$[P(X)] = \begin{bmatrix} P(x_1) & P(x_2) & \dots & P(x_m) \end{bmatrix}$$

- The output probabilities $P(Y)$ are represented by the row matrix

$$[P(Y)] = \begin{bmatrix} P(y_1) & P(y_2) & \dots & P(y_n) \end{bmatrix}$$

- We can compute $[P(Y)]$ by the following formula:

$$[P(Y)] = [P(X)] \times [P(Y|X)]$$

- (Note: Be mindful of the dimensions of each matrix).

Channel Matrix

- Suppose for our Binary DMC that the input probabilities were given by $[P(X)] = [0.5 \ 0.5]$.
- Compute $[P(Y)]$, given the channel matrix given in previous slides.

$$[P(Y)] = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \times \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- Solving

$$[P(Y)] = \begin{bmatrix} (0.5 \times 0.9) + (0.5 \times 0.2) & (0.5 \times 0.1) + (0.5 \times 0.8) \end{bmatrix}$$

- Simplifying

$$[P(Y)] = \begin{bmatrix} 0.55 & 0.45 \end{bmatrix}$$

Channel Matrix

- Let $[P(X)]$ is presented as a diagonal matrix , i.e.

$$[P(X)]_d = \begin{bmatrix} P(x_1) & 0 & \dots & 0 \\ 0 & P(x_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & P(x_m) \end{bmatrix}$$

- The **joint probability matrix** $[P(X, Y)]$ can be computed as $[P(X, Y)] = [P(X)]_d \times [P(Y|X)]$

Channel Matrix

- For the Binary DMC described in the previous example, compute the joint probability matrix.
- Diagonalize the input probabilities for X .

$$[P(X)]_d = \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$$

- Simplifying

$$[P(X, Y)] = \begin{bmatrix} (0.5 \times 0.9) + (0 \times 0.2) & (0.5 \times 0.1) + (0 \times 0.8) \\ (0 \times 0.9) + (0.5 \times 0.2) & (0 \times 0.1) + (0.5 \times 0.8) \end{bmatrix}$$

- Solving

$$[P(X, Y)] = \begin{bmatrix} 0.45 & 0.05 \\ 0.1 & 0.4 \end{bmatrix}$$

Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, the channel of Fig. 10-1 is often specified by the matrix of transition probabilities $[P(Y|X)]$, given by

The matrix $[P(Y|X)]$ is called the channel matrix. Since each input to the channel results in some output, each row of the channel matrix must sum to unity. that is,

Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, the channel from the previous slide is specified by the matrix of transition probabilities $[P(Y|X)]$, given by

$$P(Y|X) = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) & \dots & P(y_n|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & \dots & P(y_n|x_2) \\ \dots & \dots & \dots & \dots \\ P(y_1|x_m) & P(y_2|x_m) & \dots & P(y_n|x_n) \end{bmatrix} \quad (1)$$

Channel Matrix

The matrix $[P(Y|X)]$ is called the channel matrix. Since each input to the channel results in some output, each row of the channel matrix must sum to unity (i.e. all rows must add up to 1. This condition is not necessary for columns).

Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, a channel is specified by the matrix of transition probabilities $[P(Y|X)]$, given by

$$[P(Y|X)] = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) & \dots & P(y_n|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & \dots & P(y_n|x_2) \\ \dots & \dots & \dots & \dots \\ P(y_1|x_m) & P(y_2|x_m) & \dots & P(y_n|x_n) \end{bmatrix}$$

The matrix $[P(Y|X)]$ is called the *channel matrix*.

Channel Matrix

- Since each input to the channel results in some output, each row of the channel matrix must sum to unity (i.e. all rows must add up to 1. This condition is not necessary for columns).
- For the binary DMC presented previously, the channel matrix is

$$[P(Y|X)] = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- (Remark: This is not a binary symmetric channel)

Channel Matrix

- The input probabilities $P(X)$ are represented by the row matrix

$$[P(X)] = \begin{bmatrix} P(x_1) & P(x_2) & \dots & P(x_m) \end{bmatrix}$$

- The input probabilities $P(Y)$ are represented by the row matrix

$$[P(Y)] = \begin{bmatrix} P(y_1) & P(y_2) & \dots & P(y_n) \end{bmatrix}$$

- We can compute $[P(Y)]$ by the following formula:

$$[P(Y)] = [P(X)] \times [P(Y|X)]$$

- (Note: Be mindful of the dimensions of each matrix).

Channel Matrix

- Suppose for our Binary DMC that the input probabilities were given by $[P(X)] = [0.5 \ 0.5]$.
- Compute $[P(Y)]$, given the channel matrix given in previous slides.

$$[P(Y)] = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \times \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- Solving

$$[P(Y)] = \begin{bmatrix} (0.5 \times 0.9) + (0.5 \times 0.2) & (0.5 \times 0.1) + (0.5 \times 0.8) \end{bmatrix}$$

- Simplifying

$$[P(Y)] = \begin{bmatrix} 0.55 & 0.45 \end{bmatrix}$$

Channel Matrix

- Let $[P(X)]$ is presented as a diagonal matrix , i.e.

$$[P(X)]_d = \begin{bmatrix} P(x_1) & 0 & \dots & 0 \\ 0 & P(x_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & P(x_m) \end{bmatrix}$$

- The **joint probability matrix** $[P(X, Y)]$ can be computed as $[P(X, Y)] = [P(X)]_d \times [P(Y|X)]$

Channel Matrix

- For the Binary DMC described in the previous example, compute the joint probability matrix.
- Diagonalize the input probabilities for X .

$$[P(X)]_d = \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$$

- Simplifying

$$[P(X, Y)] = \begin{bmatrix} (0.5 \times 0.9) + (0 \times 0.2) & (0.5 \times 0.1) + (0 \times 0.8) \\ (0 \times 0.9) + (0.5 \times 0.2) & (0 \times 0.1) + (0.5 \times 0.8) \end{bmatrix}$$

- Solving

$$[P(X, Y)] = \begin{bmatrix} 0.45 & 0.05 \\ 0.1 & 0.4 \end{bmatrix}$$

Types of Channel

1. Lossless Channel:

A channel described by a channel matrix with only one non-zero element in each column is called a lossless channel.

$$[P(Y|X)] = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(Drawn in overhead):

It can be shown that in the lossless channel, no source information is lost in transmission.

Types of Channel

2. Deterministic Channel:

A channel described by a channel matrix with only one non-zero element in each row is called a deterministic channel.

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(Drawn in overhead):

Note that since each row has only one non-zero element, this element must be 1. When a given source symbol is sent in a deterministic channel, it is clear which output symbol would be received.

Types of Channel

3. Noiseless Channel:

A channel is said to be *noiseless* if it is both lossless and deterministic. The channel matrix is the identity matrix: only one element in each row and each column, and each element is necessarily 1.

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

(Drawn in overhead):

Note that the input and output alphabets have the same size , i.e. $m = n$.

Types of Channel

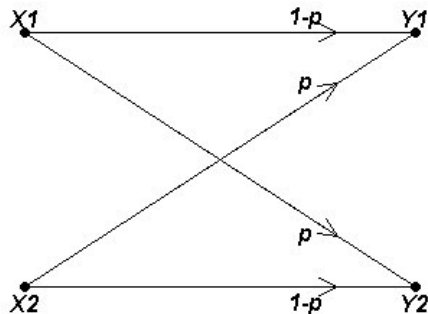
4. Binary Symmetric Channel:

The binary symmetric channel is defined by the following channel diagram (next slide) and the channel matrix is given by

$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

The channel has two inputs and two outputs ($x_1 = 0, x_2 = 1$) and two outputs ($x_1 = 0, x_2 = 1$). The channel is symmetric because the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 was sent. This common probability is denoted p .

Binary Symmetric Channels



Types of Channel

1. Lossless Channel:

A channel described by a channel matrix with only one non-zero element in each column is called a lossless channel.

$$[P(Y|X)] = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(Drawn in overhead):

It can be shown that in the lossless channel, no source information is lost in transmission.

Types of Channel

2. Deterministic Channel:

A channel described by a channel matrix with only one non-zero element in each row is called a deterministic channel.

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(Drawn in overhead):

Note that since each row has only one non-zero element, this element must be 1. When a given source symbol is sent in a deterministic channel, it is clear which output symbol would be received.

Types of Channel

3. Noiseless Channel:

A channel is said to be *noiseless* if it is both lossless and deterministic. The channel matrix is the identity matrix: only one element in each row and each column, and each element is necessarily 1.

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

(Drawn in overhead):

Note that the input and output alphabets have the same size , i.e. $m = n$.

Types of Channel

4. Binary Symmetric Channel:

The binary symmetric channel is defined by the following channel diagram (next slide) and the channel matrix is given by

$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

The channel has two inputs and two outputs ($x_1 = 0, x_2 = 1$) and two outputs ($x_1 = 0, x_2 = 1$). The channel is symmetric because the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 was sent. This common probability is denoted p .

Binary Symmetric Channels

