Statistics for Computing

MA4413 Lecture 11A

Kevin O'Brien

kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

Autumn 2013

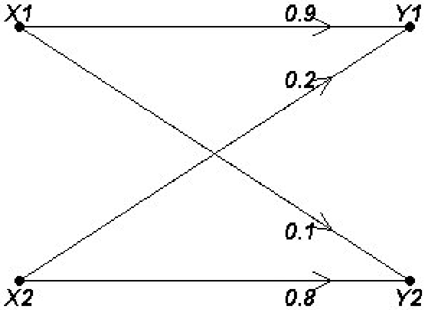
Discrete memoryless channel (from last lecture)

- For a DMC with “m” inputs and “n” outputs, the input X consists of input symbols $x_1, x_2, \ldots x_m$.
- The probabilities of these source symbols $P(x_i)$ are assumed to be known.
- The output Y consists of output symbols $\{y_1, y_2, \ldots, y_n\}$
- 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



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 the same as 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.

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

- 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.

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

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

Types of Channel

1. Lossless Channel:

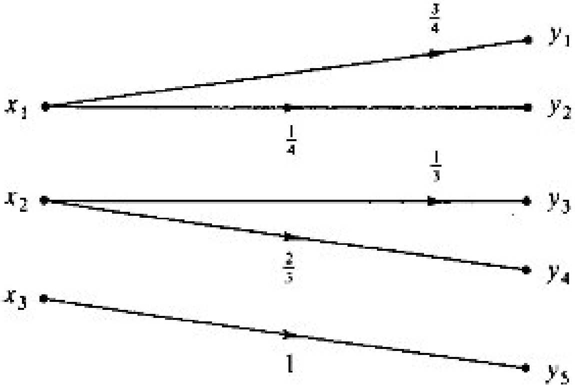


Figure:

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.

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

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

2. Deterministic Channel:

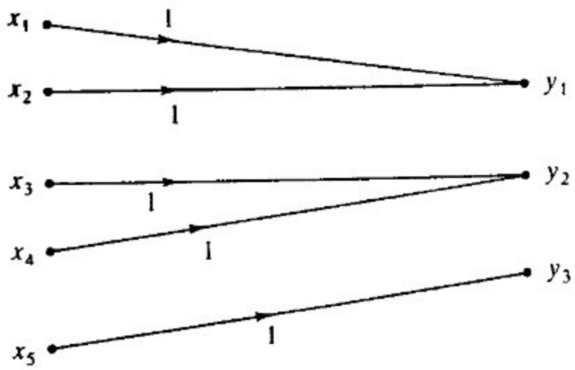


Figure:

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.

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

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

Types of Channel

3. Noiseless Channel:

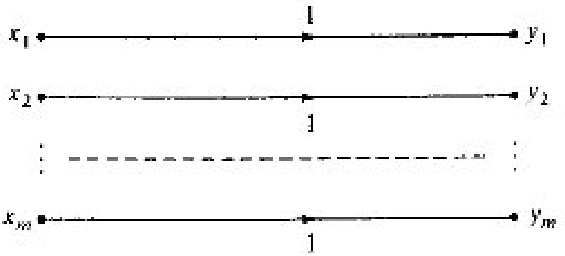


Figure:

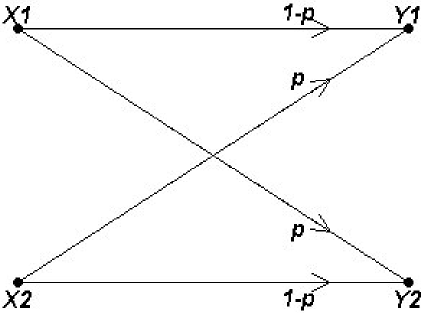
Types of Channel

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



Mutual Information

Mutual information is one of many quantities that measures how much one random variables gives about another. It is a dimensionless quantity. Mutual Informaiton can be thought of as the reduction in uncertainty about one random variable given knowledge of another.

- High mutual information indicates a large reduction in uncertainty,
- low mutual information indicates a small reduction,
- zero mutual information between two random variables means the variables are independent.

Efficient communication systems have high mutual information.

Mutual Information

Joint Entropies:
Using the input probabilities $P(x_i)$, output probabilities $P(y_i)$, transition probabilities $P(y_i|x_i)$, and joint probabilities $P(x_i, y_j)$, we can define the following various entropy functions for a channel with m inputs and n outputs:

- $H(X) = -\sum_{i=1}^{m} P(x_i)\log_2 P(x_i)$
- $H(Y) = -\sum_{j=1}^{n} P(y_j)\log_2 P(y_j)$
- $H(X,Y) = -\sum_{j=1}^{n} \sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i, y_j)$

Mutual Information: Joint Entropy

These entropies can be interpreted as follows:

- $H(X)$ is the average uncertainty of the channel input, and $H(Y)$ is the average uncertainty of the channel output.
- The joint entropy $H(X,Y)$ is the average uncertainty of the communication channel as a whole.

Mutual Information: Conditional Entropy

- The conditional entropy $H(X|Y)$ is a measure of the average uncertainty remaining about the channel input after the channel output has been observed.
- This is sometimes called the equivocation of X with respect to Y.
- The conditional entropy $H(Y|X)$ is the average uncertainty of the channel output given that X was transmitted.

- $H(X|Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i|y_j)$
- $H(Y|X) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(y_j|x_i)$

Mutual Information : Useful Identities

Two useful relationships among the types of entropies are

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

(Remark : compare to identities in probability theory)

Mutual Information : Useful Identities

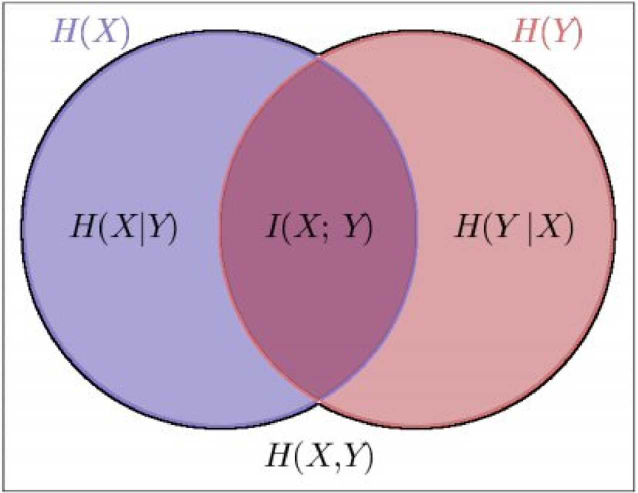


Figure:

Mutual Information

The mutual information $I(X;Y)$ of a channel is defined by

$$I(X;Y) = H(X) - H(X|Y) \text{ (b/symbol)}$$

Alternatively we can define it as either

$$I(X;Y) = H(Y) - H(Y|X) \text{ (b/symbol)}$$

or as

$$I(X;Y) = H(Y) + H(Y) - H(X,Y) \text{ (b/symbol)}$$

Remark: The mutual information is the reduction of entropy of X when Y is known.

Entropy: some remarks

- recall: Entropy is the uncertainty of a single random variable.
- We can define ***conditional entropy*** $H(X|Y)$, which is the entropy of a random variable conditional on the knowledge of another random variable.
- The reduction in uncertainty due to another random variable is called the ***mutual information***.

Entropies: Example

- The input source to a noisy communication channel is a random variable X over the four symbols $\{a, b, c, d\}$.
- The output from this channel is a random variable Y over these same four symbols.

Entropies: Example

The joint distribution of these two random variables is as follows:

	x=a	x=b	x=c	x=d
y=a	1/8	1/16	1/16	1/4
y=b	1/16	1/8	1/16	0
y=c	1/32	1/32	1/16	0
y=d	1/32	1/32	1/16	0

Entropies: Example

- Write down the marginal distribution for X and compute the marginal entropy $H(X)$.
- Write down the marginal distribution for Y and compute the marginal entropy $H(Y)$.
- (next class) What is the joint entropy $H(X,Y)$ of the two random variables?
- (next class) What is the conditional entropy $H(Y|X)$?
- (next class) What is the conditional entropy $H(X|Y)$?
- (next class) What is the mutual information $I(X;Y)$ between the two random variables?

Entropies: Example

The marginal distribution of these two random variables is as follows:

	x=a	x=b	x=c	x=d	P(Y)
y=a	1/8	1/16	1/16	1/4	0.50
y=b	1/16	1/8	1/16	0	0.25
y=c	1/32	1/32	1/16	0	0.125
y=d	1/32	1/32	1/16	0	0.125
P(X)	0.25	0.25	0.25	0.25	

Entropies: Example

- H(X) , the entropy of X, is computed as

$$H(X) = -\sum P(x_i)\log_2 P(x_i)$$

- $H(X) = (-0.25 \times -2) + (-0.25 \times -2) + (-0.25 \times -2) + (-0.25 \times -2)$
- $H(X) = 2\text{b}$

- H(X) , the entropy of Y, is computed as

$$H(Y) = -\sum P(y_j)\log_2 P(y_j)$$

- $H(Y) = (-0.5 \times -1) + (-0.25 \times -2) + (-0.125 \times -3) + (-0.125 \times -3)$
- $H(Y) = 1.75\text{b}$