ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Multimedia Technology

Section # 4: Information Theory

Instructor: George Xylomenos

Department: Informatics

Contents

- Channels
- Information
- Applications of information
- Entropy
- Applications of entropy

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Channels

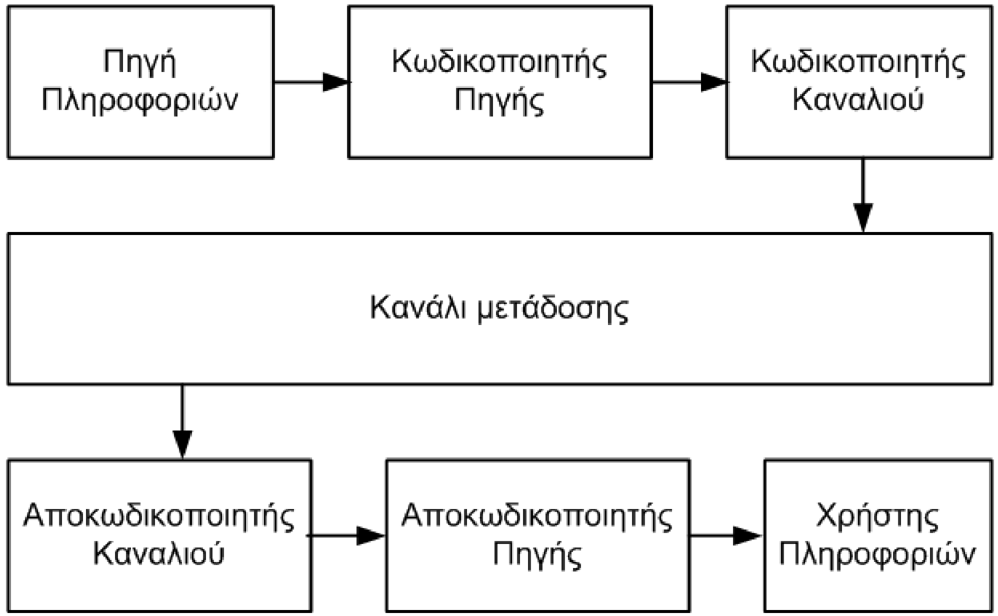
Class: Multimedia Technology, **Section # 4:** Information Theory

Instructor: George Xylomenos, **Department:** Informatics

Channels (1 of 4)

- Transmission of digital information
  - Discrete symbols from a discrete alphabet
  - Source / Channel / User
- Source coding
  - Reduces the bits to transmit (compression)
  - Based on the nature of information
- Channel coding
  - Improves channel reliability
  - Based on the nature of the channel

Channels (2 of 4)



- Discrete Memoryless Channel (DMC)
  - Independent transmission of symbols (M)
  - Discrete symbols (D)

Channels (3 of 4)

- Channels can be different things
  - Transmission channel: WiFi link
    - Uses OFDM to encode source symbols
    - Goal: guard against interference
  - Storage channel: CD
    - Uses RS coding to encode source symbols
    - Goal: guard against disc errors or damage
  - What is the capacity of a channel?

Channels (4 of 4)

- Shannon-Hartley theorem

$$C = B\ log_2(1 + \frac{S}{N})$$

  - C: Capacity in bits per second
  - B: Bandwidth in Hz
  - S: Signal power (W or V^2)
  - N: Noise/interference power (W or V^2)
    - S/N: Signal to Noise Ratio (SNR)

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Information

Class: Multimedia Technology, **Section # 4:** Information Theory

Instructor: George Xylomenos, **Department:** Informatics

Measuring information (1 of 2)

- How can we quantify information?
  - How much information is on a page?
    - Is a printed page the same as a blank page?
  - How much information is in an image?
    - Is it a vector image or a bit map?
  - What are the elements of information?
    - Pixels, lines, letters, logos?
  - How much can I compress information?

Measuring information (2 of 2)

- How much information does a channel carry?
  - The maximum is what Shannon-Hartley says
  - But this is not a recipe for how to transmit
  - In practice, channels are imperfect
    - The transmitter sends symbol X
    - The receiver gets symbol Y
    - Y may not be the same as X (e.g., a bit was flipped)
    - How much information did the channel carry?

Mutual information (1 of 3)

- Assume two discrete random variables X and Y
  - x_i, i=1, 2, …, n and y_j, j=1, 2, …, m
- If X and Y are statistically independent
  - $Y=y_j$ does not provide any information on $X=x_i$
- If X and Y are fully dependent
  - $Y=y_j$ assures as that $X=x_i$
- What information does Y provide for X?
  - When I observe Y, what do I learn about X?
  - How can I quantify this information?

Mutual information (2 of 3)

- Proportional to conditional probability

$$P(X = x_i | Y = y_j) = P(x_i | y_j)$$

- Normalized to the magnitude of $X=x_i$

$$P(X = x_i) = P(x_i)$$

- $I(x_i;y_j)$: mutual information between x_i and y_j

$$I(x_i; y_j) = \log \frac{P(x_i | y_j)}{P(x_i)}$$

Mutual information (3 of 3)

- Mutual information is symmetric

$$\frac{P(x_i|y_j)}{P(x_i)} = \frac{P(x_i|y_j)P(y_j)}{P(x_i)P(y_j)} = \frac{P(x_i, y_j)}{P(x_i)P(y_j)} = \frac{P(y_j|x_i)P(x_i)}{P(y_j)P(x_i)} = \frac{P(y_j|x_i)}{P(y_j)}$$

$$I(x_i; y_j) = \log\frac{P(x_i|y_j)}{P(x_i)} = \log\frac{P(y_j|x_i)}{P(y_j)} = I(y_j; x_i)$$

- X and Y statistically independent: $I(x_i;y_j)=0$

$$I(x_i; y_j) = \log\frac{P(x_i|y_j)}{P(x_i)} = \log\frac{P(x_i)}{P(x_i)} = 0$$

- X and Y fully dependent

$$I\left(x_i; y_j\right) = \log\frac{P(x_i|y_j)}{P(x_i)} = \log\frac{1}{P(x_i)} = -\log P\,(x_i)$$

(Intrinsic) Information

- What is full dependence between X and Y?
  - The channel between them is perfect
  - It carries the full information of the source
- (Intrinsic) Information of x_i

$$I(x_i) = \log \frac{1}{P(x_i)} = -\log P\,(x_i)$$

  - Non-negative
    - Log of a number <= 1 is always negative
  - Measured in binary digits, or bits ($\log_2 x$ or lg x)

Conditional information (1 of 2)

- Definition of conditional information
  - $X=x_i$ given that $Y=y_j$

$$I(x_i|y_j) = \log\frac{1}{P(x_i|y_j)} = -\log P\,(x_i|y_j)$$

- Mutual information $I(x_i;y_j)$
  - Information that $Y=y_j$ provides for $X=x_i$
- Conditional information $I(x_i|y_j)$
  - Intrinsic information of $X=x_i$ …
  - … when $Y=y_j$ occurs

Conditional information (2 of 2)

- Conditional information: non-negative
  - Same reason as for intrinsic information
- All these definitions are connected

$$I(x_i; y_j) = \log \frac{P(x_i|y_j)}{P(x_i)} = \log P\,(x_i|y_j) - \log P\,(x_i) = I(x_i) - I(x_i|y_j)$$

  - $I(x_i;y_j)>0$ when $I(x_i)>I(x_i|y_j)$
  - $I(x_i;y_j)<0$ when $I(x_i)<I(x_i|y_j)$

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Applications of information

Class: Multimedia Technology, **Section # 4:** Information Theory

Instructor: George Xylomenos, **Department:** Informatics

Application to sources (1 of 2)

- Assume a binary source (e.g., it produces bits)
  - If P(0)=P(1)=1/2

$$I(x_i) = -\log_2 P(x_i) = -\log_2 \frac{1}{2} = 1$$

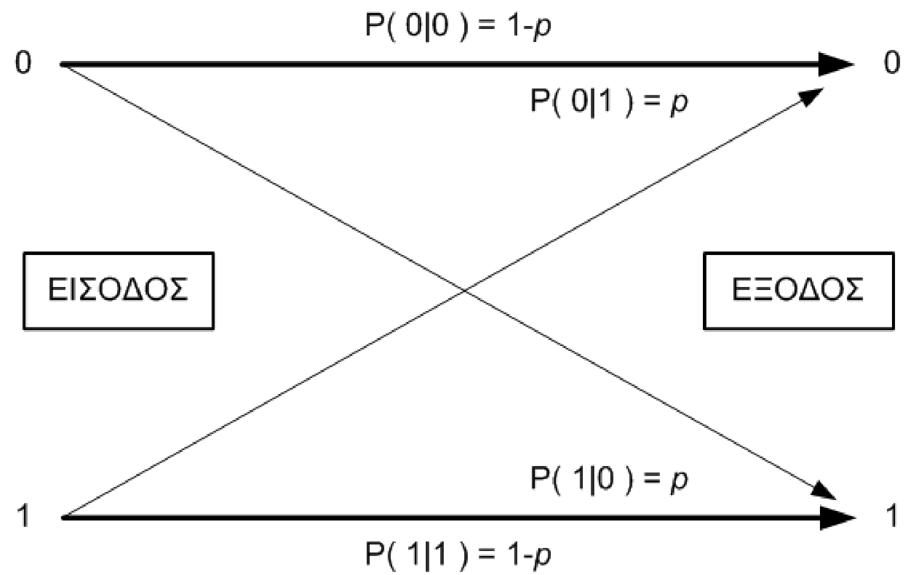
- Memoryless binary source
  - k consecutive symbols
  - $M=2^k$ different sequences of symbols
  - $P(x_i') = 1/M = 2^{-k}$

$$I(x_i') = -\log_2 2^{-k} = k$$

Application to sources (2 of 2)

- Information is a logarithmic measure
  - Logarithms turn multiplication to addition
    - log ab = log a + log b
  - What is the information of a sequence of events?
  - It is the sum of the information of each event
    - 1 random binary symbol -> 1 bit
    - k random binary symbols -> k bits
  - The base 2 logarithm turns this to bits

Application to transmission (1 of 3)



- Assume a binary DMC (transmitting bits)
  - X: input, transmitted signal
  - Y: output, received signal
  - Output differs from input with probability p
  - Assume that P(X=0)=P(X=1)=1/2 (random source)

Application to transmission (2 of 3)

- Probability of output being 0 or 1

$$P(Y = 1) = P(Y = 1|X = 0)P(X = 0) + P(Y = 1|X = 1)P(X = 1) = \frac{1}{2}(p + 1 - p) = \frac{1}{2}$$

$$P(Y = 0) = P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = 1)P(X = 1) = \frac{1}{2}(1 - p + p) = \frac{1}{2}$$

- Mutual information

$$I(y_0; x_0) = I(0; 0) = \log_2 \frac{P(Y = 0|X = 0)}{P(Y = 0)} = \log_2 2\,(1 - p)$$

$$I(y_0; x_1) = I(0; 1) = \log_2 \frac{P(Y = 0|X = 1)}{P(Y = 0)} = \log_2 2\,p$$

Application to transmission (3 of 3)

- Noiseless channel
  - p=0: you always get what you sent
  - Therefore, I(0;0)=1
- Noisy channel
  - p=1/2, therefore I(0;0)=I(0;1)=0
  - p=1/4, therefore I(0;0)=0,587 and I(0;1)=-1
  - Note that p=1/2 is the worst case!
    - It makes the channel completely random

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Entropy

Class: Multimedia Technology, **Section # 4:** Information Theory

Instructor: George Xylomenos, **Department:** Informatics

Information and entropy (1 of 4)

- Assume an information source
  - Produces a set of symbols
  - We know the information for each symbol
- How can we characterize the source?
  - We weigh all symbols…
  - …based on their probability
  - Essentially, a weighted mean of information

Information and entropy (2 of 4)

- Mean mutual information of X and Y

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) I(x_i; y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log \frac{P(x_i|y_j)}{P(x_i)}$$

- $P(x_i,y_j)$ and $I(x_i;y_j)$ are symmetric
  - I(X;Y)= I(Y;X)
- X and Y statistically independent
  - $P(x_i|y_j)=P(x_i)$, άρα I(X;Y)=0

Information and entropy (3 of 4)

- Mean (intrinsic) information X

$$H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) \log P\,(x_i)$$

- Values of X: symbols of an alphabet
  - H(X): entropy of the source
- Source with random behavior
  - $P(x_i)=1/n$

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n$$

  - Requires $\log_2 n$ bits per symbol

Information and entropy (4 of 4)

- Maximum entropy
  - All symbols are equally probable
- Conditional entropy

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

- All these definitions can be combined

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

Applications of entropy

Class: Multimedia Technology, **Section # 4:** Information Theory

Instructor: George Xylomenos, **Department:** Informatics

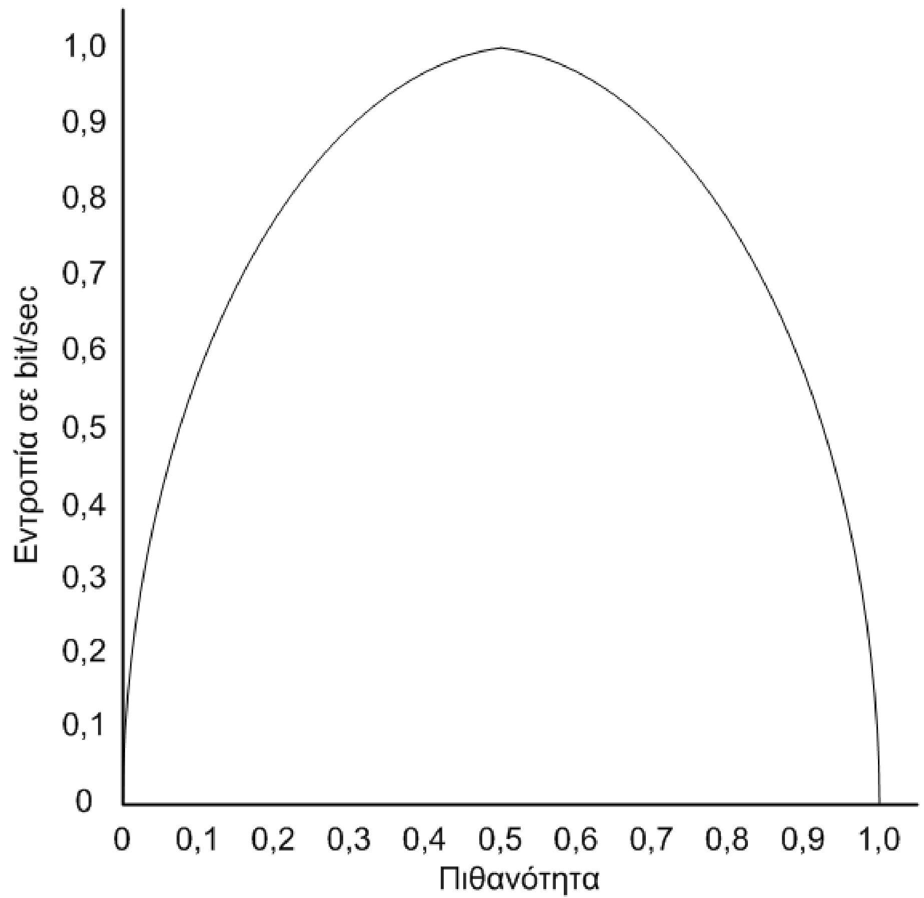
Behavior of entropy (1 of 3)

- Discrete memoryless source
  - P(0)=q, P(1)=1-q
- Source entropy

$$H(X) = -P(0)\log P(0) - P(1)\log P(1) =$$

$$= -q\log q - (1-q)\log(1-q)$$

- Binary entropy function
  - Depends on q
  - When q=1-q=1/2, H(X)=1
  - When q=0 ή q=1, H(X)=0



Behavior of entropy (2 of 3)

- Discrete memoryless channel (DMC)
- Entropy of source X

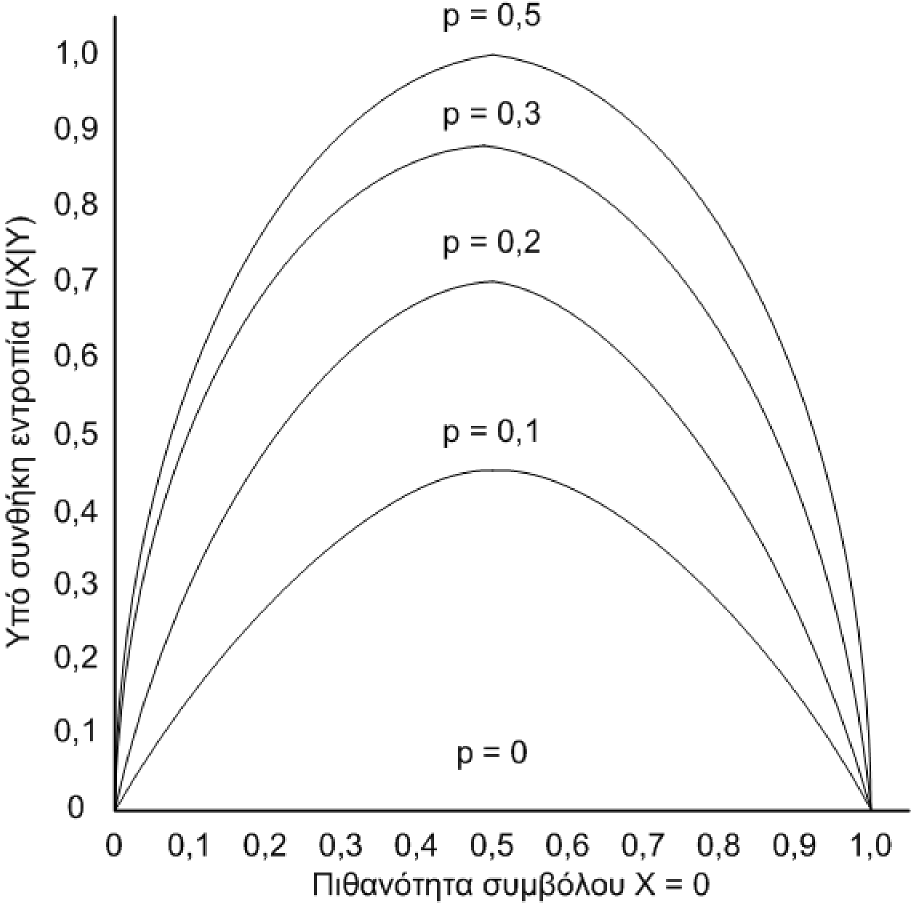
$$H(X) \equiv H(q) = -q \log q - (1-q) \log(1-q)$$

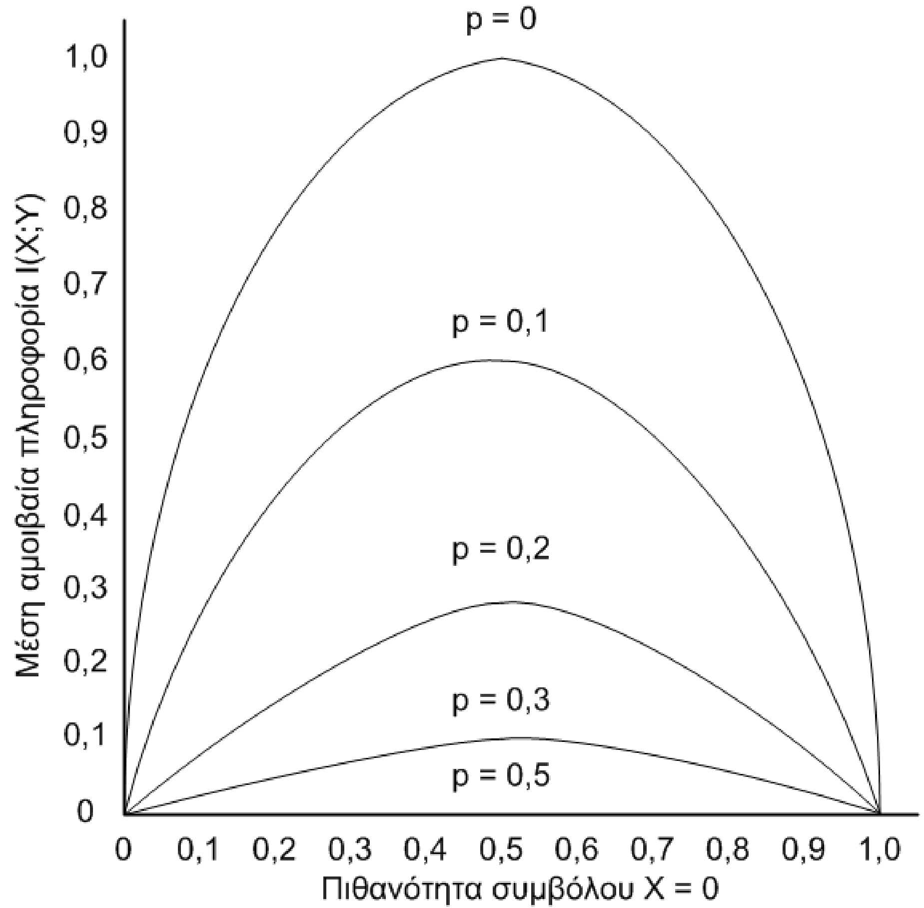
- Mean mutual information I(X;Y)

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

  - Maximized when q=1-q=1/2 for every p
  - p=0: maximum, p=1/2: minimum
- Conditional entropy H(X|Y)
  - The opposite behavior from I(X;Y)
  - p=1/2: maximum, p=0: minimum

Behavior of entropy (3 of 3)





- Mean mutual information and conditional entropy
  - Complementary to each other

Applications (1 of 4)

- Assume a media source
  - Produces w symbols q_i, i=1, 2, …, w
  - Probability of symbol q_i is $P(q_i)$
- Simple symbol representation
  - Fixed length binary sequences
  - Need $\log_2 w$ bits per symbol for w symbols
  - Example: US ASCII
  - 128 characters encoded with 7 bits

Applications (2 of 4)

- Information
  - $I(q_i)=-\log_2 P(q_i)$ bits
- Variable length representation
  - $I(q_i)$ bits for symbol q_i
  - Differentiates common and rare symbols
    - Fewer bits for common ones
    - More bits for rare ones
  - This is the most economical encoding
    - But: this is a limit, not an encoding method

Applications (3 of 4)

- Source entropy H(X)
  - Average number of information bits/symbol
- Efficiency of a coding scheme
  - R: Average number of coded bits/symbol
  - Efficiency: H(X)/R
  - Goal: get as close to the optimal as possible
- Note: optimal under specific assumptions
  - Lossless, per symbol encoding

Applications (4 of 4)

- Transmission over a channel
  - X: input, Y: output
- Conditional entropy H(X|Y)
  - Mean input information, when we know the output
- Entropy H(X)
  - Mean input information, regardless of output
- Mean mutual information I(X;Y)
  - $I(X;Y)=H(X)-H(X|Y)$
  - Difference of information before – after transmission

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

End of Section #4

Class: Multimedia Technology, **Section # 4:** Information Theory

Instructor: George Xylomenos, **Department:** Informatics