

ECE 515

Information Theory

Joint Entropy, Equivocation and Mutual
Information I

Entropy

$$H(X) = - \sum_{i=1}^N p(x_i) \log_b p(x_i)$$

Joint Entropy

$$H(XY) = - \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log_b p(x_i, y_j)$$

Conditional Entropy

$$H(X|Y) = - \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log_b p(x_i|y_j)$$

Chain Rule

$$H(XY) = H(X) + H(Y|X)$$

$$H(X_1, \dots, X_N) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \dots + H(X_N|X_1, \dots, X_{N-1})$$

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

Example

- X, Y, Z binary RVs
 - $x_1 = 0, x_2 = 1, y_1 = 0, y_2 = 1, z_1 = 0, z_2 = 1$
- Four equally likely vectors (probability of each $\frac{1}{4}$)
 - $[x, y, z] = [0, 0, 0]$
 - $[0, 1, 0]$
 - $[1, 0, 0]$
 - $[1, 0, 1]$
- Find $H(XYZ)$ using
$$H(XYZ) = H(X) + H(Y|X) + H(Z|XY)$$

Example

$$H(X) = - \sum_{i=1}^2 p(x_i) \log_2 p(x_i)$$

$$H(Y|X) = - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log_2 p(y_j|x_i)$$

$$H(Z|XY) = - \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 p(x_i, y_j, z_k) \log_2 p(z_k|x_i, y_j)$$

Example

$$p(x_1) = p(x_2) = \frac{1}{2}$$

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \text{ bit}$$

$$p(x_i, y_j)$$

$$p(0,0) = \frac{1}{4}$$

$$p(1,1) = 0$$

$$p(0,1) = \frac{1}{4}$$

$$p(1,0) = \frac{1}{2}$$

$$p(y_j | x_i)$$

$$p(0 | 0) = \frac{1}{2}$$

$$p(1 | 1) = 0$$

$$p(1 | 0) = \frac{1}{2}$$

$$p(0 | 1) = 1$$

$$\begin{aligned} H(Y | X) &= -\frac{1}{4}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{2} - \frac{1}{2}\log_2 1 \\ &= \frac{1}{2}\log_2 2 = \frac{1}{2} \text{ bit} \end{aligned}$$

Example

$$p(x_i, y_j, z_k)$$

$$p(0,0,0) = \frac{1}{4}$$

$$p(0,1,0) = \frac{1}{4}$$

$$p(1,0,0) = \frac{1}{4}$$

$$p(1,0,1) = \frac{1}{4}$$

$$p(z_k | x_i, y_j)$$

$$p(0 | 0,0) = 1$$

$$p(0 | 0,1) = 1$$

$$p(0 | 1,0) = \frac{1}{2}$$

$$p(1 | 1,0) = \frac{1}{2}$$

$$\begin{aligned} H(Z | XY) &= -\frac{1}{4}\log_2 1 - \frac{1}{4}\log_2 1 - \frac{1}{4}\log_2 \frac{1}{2} - \frac{1}{4}\log_2 \frac{1}{2} \\ &= \frac{1}{2}\log_2 2 = \frac{1}{2} \text{ bit} \end{aligned}$$

Example

- $H(XYZ) = H(X) + H(Y|X) + H(Z|XY)$
- $H(X) = 1$ bit
- $H(Y|X) = \frac{1}{2}$ bit
- $H(Z|XY) = \frac{1}{2}$ bit
- $H(XYZ) = 1 + \frac{1}{2} + \frac{1}{2} = 2$ bits

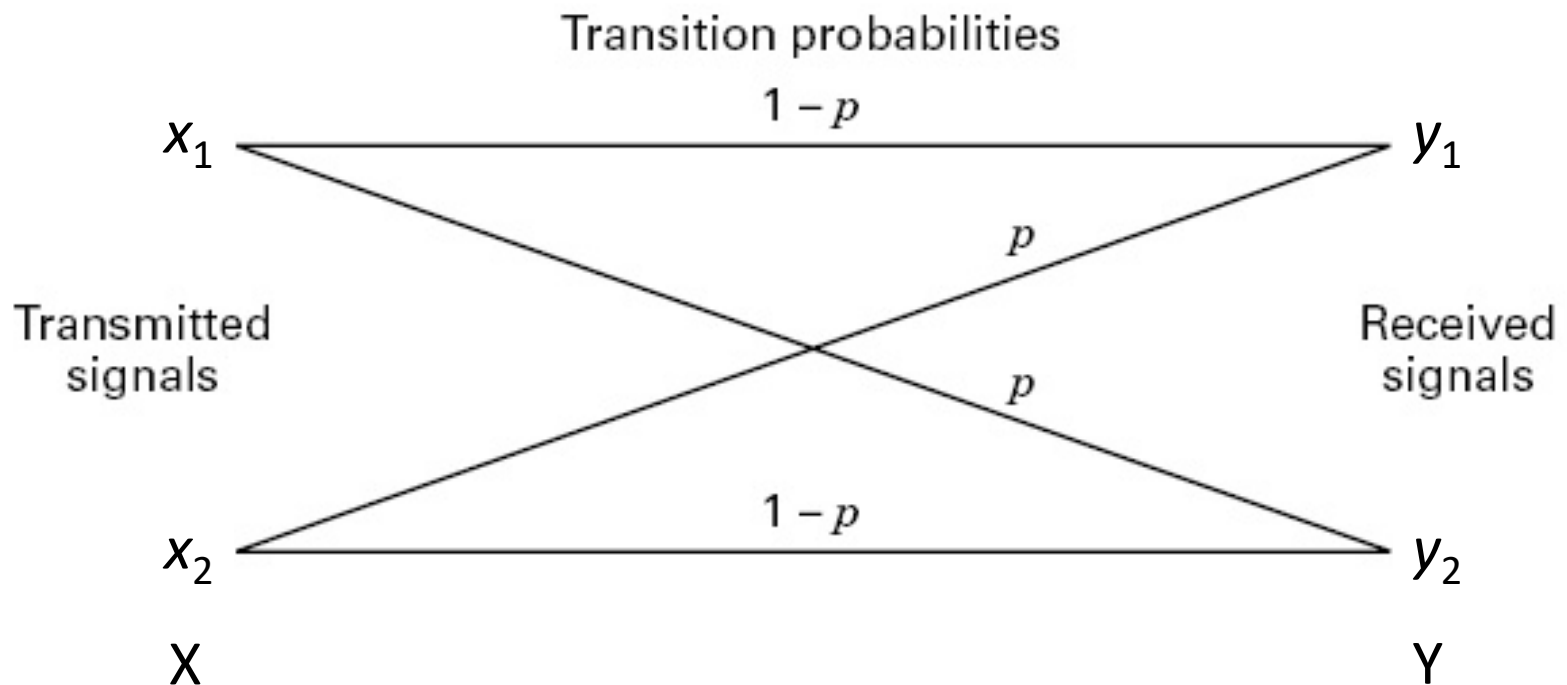
- $p(y_1) = \frac{3}{4}$ $p(y_2) = \frac{1}{4}$
- $H(Y) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = .811$ bit $> H(Y|X)$

Information Channels

- An information channel is described by an
- Input random variable X
- Output random variable Y
- Set of conditional probabilities $p(y_j | x_i)$



Binary Symmetric Channel



Information Channels

- An important insight is that an information channel does not have to involve communication

$X \rightarrow \text{physical/statistical process} \rightarrow Y$

where

X is something we want to know

the middle process transforms or obscures information

Y is something we can observe

- Information theory asks how much knowledge about X can be obtained from Y
- Applications include medical diagnosis, astronomy, seismology, archaeology, genetics, forensics, and machine learning

Medical Diagnosis

- Suppose

X = patient's actual disease

and

Y = result of a diagnostic test

- The biological and measurement processes form an information channel
- For example
Disease $\rightarrow$ biological changes $\rightarrow$ blood test $\rightarrow$ test result
- The test is not perfectly reliable, so Y does not completely reveal X
- One can ask:
How much information does the test result provide about the patient's actual disease?

Astronomy

- Consider trying to determine the temperature of a distant star
- The star's physical properties are the source

X = star's temperature

- The emitted radiation travels through space and interacts with the telescope and detector

$X \rightarrow \text{physical propagation} \rightarrow Y$

- The astronomer observes a spectrum Y and tries to infer X
- This is an information channel problem
- A useful question is:

How much information about the star's temperature is contained in its observed spectrum?

Genetics

- Genetic inheritance can be viewed this way

parental DNA → replication/recombination/mutation → offspring DNA

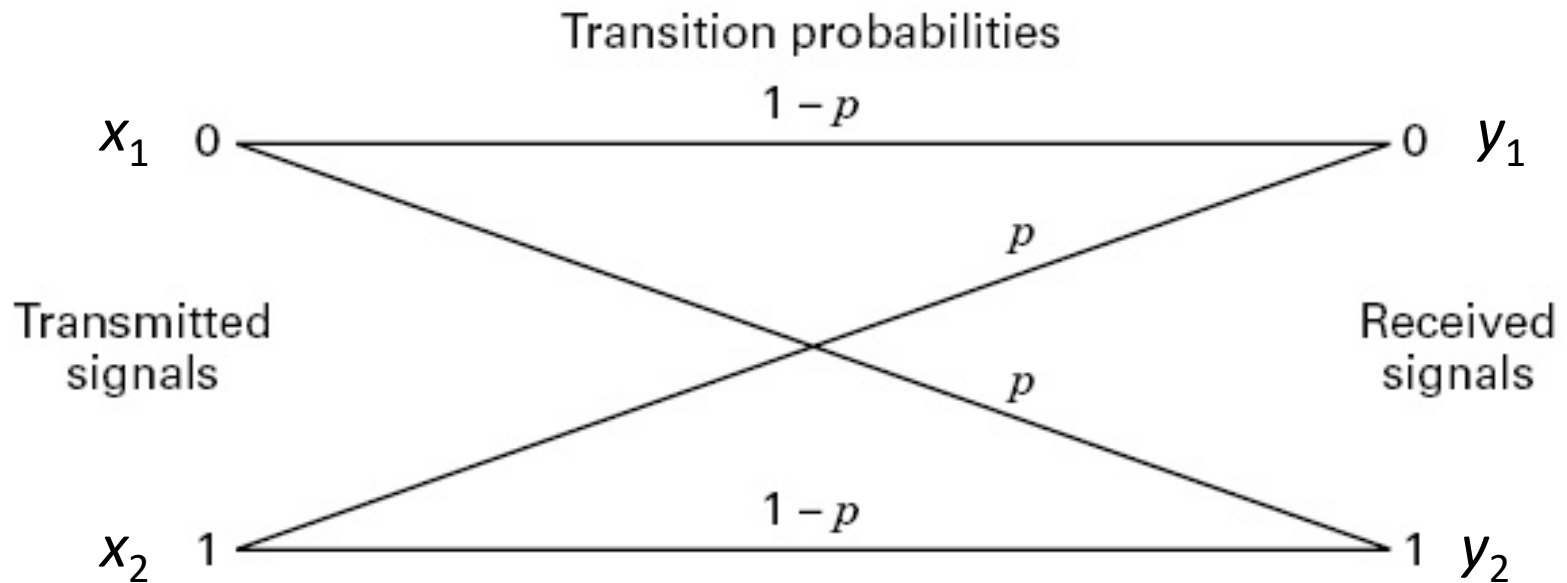
- Mutation and recombination introduce transformations between the input and output
- One could ask:
How much information about parental genetic material is retained in the offspring?
- This provides an interesting connection between **information theory** and **evolutionary biology**

Machine Learning Classification

- Suppose an image depicts one of K objects
- Let X = true object and Y = classifier output
- The transformation
 $X \rightarrow \text{image/features} \rightarrow \text{ML model} \rightarrow Y$
can be considered as an information channel
- For example, $X \in \{\text{cat}, \text{dog}\}$ and the classifier output is $Y \in \{\text{cat}, \text{dog}\}$
- One can construct a matrix representing the channel
- For example

$$P(Y | X) = \begin{pmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{pmatrix}$$

Binary Symmetric Channel



$$p_{Y|X}(0|1) = p_{Y|X}(1|0) = p$$

$$p_{Y|X}(0|0) = p_{Y|X}(1|1) = 1-p$$

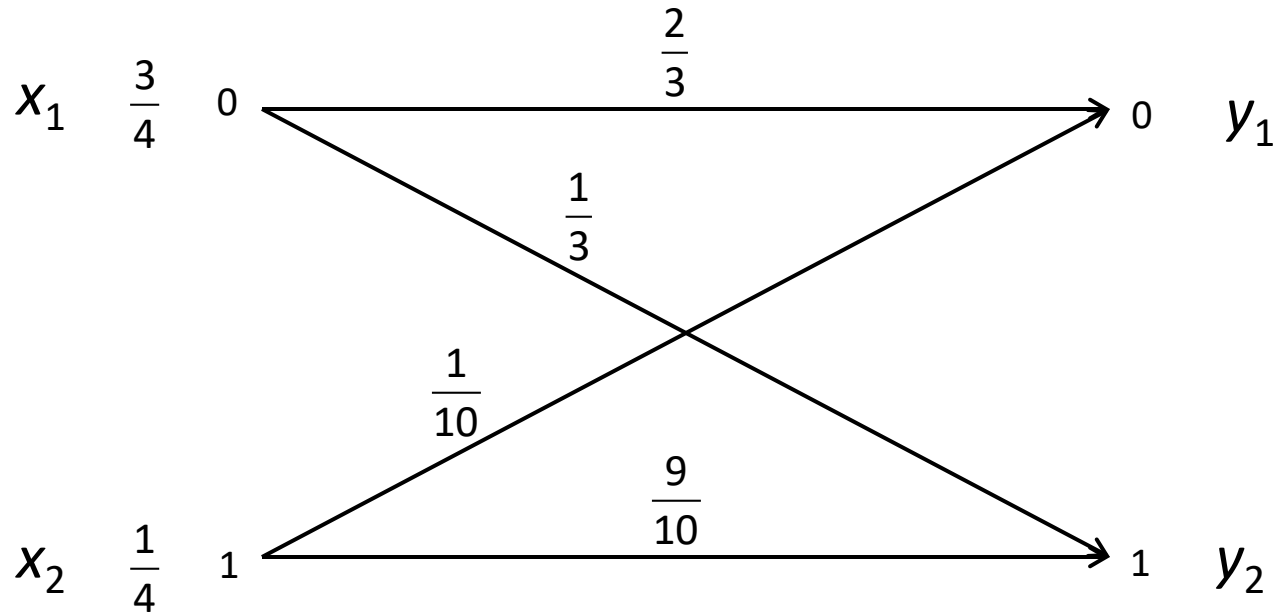
$$\begin{bmatrix} p(y_1 | x_1) & p(y_2 | x_1) \\ p(y_1 | x_2) & p(y_2 | x_2) \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

- The probabilities $p(y_j | x_i)$ are called the **forward transition probabilities**
- Using Bayes' Theorem

$$p(x_i | y_j) = \frac{p(y_j | x_i)p(x_i)}{p(y_j)}$$

- The probabilities $p(x_i | y_j)$ are called the **backward transition probabilities**

Non-symmetric Binary Channel



channel matrix $\begin{bmatrix} p(y_1 | x_1) & p(y_2 | x_1) \\ p(y_1 | x_2) & p(y_2 | x_2) \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$

Backward Transition Probabilities

$$\begin{bmatrix} p(x_1 | y_1) & p(x_2 | y_1) \\ p(x_1 | y_2) & p(x_2 | y_2) \end{bmatrix} = \begin{bmatrix} p_{x|Y}(0|0) & p_{x|Y}(1|0) \\ p_{x|Y}(0|1) & p_{x|Y}(1|1) \end{bmatrix} = \begin{bmatrix} \frac{20}{21} & \frac{1}{21} \\ \frac{10}{19} & \frac{9}{19} \end{bmatrix}$$

- $H(X|y=0) = -p(1|0)\log_2 p(1|0) - p(0|0)\log_2 p(0|0)$
 $= - (1/21)\log_2(1/21) - (20/21)\log_2(20/21)$
 $= .209 + .067 = .276 \text{ bit}$
- $H(X|y=1) = -p(1|1)\log_2 p(1|1) - p(0|1)\log_2 p(0|1)$
 $= - (9/19)\log_2(9/19) - (10/19)\log_2(10/19)$
 $= .511 + .487 = .998 \text{ bit}$
- $H(X|Y) = p(y=0) H(X|y=0) + p(y=1) H(X|y=1)$
 $= (21/40) \times (.276) + (19/40) \times (.998) = .619 \text{ bit}$

Conditional Entropy

- $H(XY) = H(X) + H(Y|X)$
- $H(XY) = H(Y) + H(X|Y)$
- If X and Y are statistically independent
 - $H(X) = H(X|Y)$
 - $H(Y) = H(Y|X)$
 - $H(XY) = H(X) + H(Y)$
- In general
 - $H(XY) \leq H(X) + H(Y)$
 - $H(X) \geq H(X|Y)$
 - $H(Y) \geq H(Y|X)$

Two Questions

- Given two random variables X and Y
 - How much information does Y give about X ?
 - How much information does X give about Y ?

Article

Interpreting Social Accounting Matrix (SAM) as an Information Channel

Mateu Sbert ^{1,*}, Shuning Chen ², Miquel Feixas ¹, Marius Vila ¹ and Amos Golan ^{3,4}

¹ Graphics and Imaging Laboratory, University of Girona, 17003 Girona, Spain; feixas@ima.udg.edu (M.F.); marius.vila@gmail.com (M.V.)

² Research Institute of Innovative Technology for the Earth, Kyoto 6190292, Japan; chensn@rite.or.jp

³ Department of Economics, American University, Washington, DC 20016, USA; agolan@american.edu

⁴ Sante Fe Institute, Albuquerque, NM 87501, USA

* Correspondence: mateu@ima.udg.edu

Received: 25 October 2020; Accepted: 24 November 2020; Published: 28 November 2020



Abstract: Information theory, and the concept of information channel, allows us to calculate the mutual information between the source (input) and the receiver (output), both represented by probability distributions over their possible states. In this paper, we use the theory behind the information channel to provide an enhanced interpretation to a Social Accounting Matrix (SAM), a square matrix whose columns and rows present the expenditure and receipt accounts of economic actors. Under our interpretation, the SAM's coefficients, which, conceptually, can be viewed as a Markov chain, can be interpreted as an information channel, allowing us to optimize the desired level of aggregation within the SAM. In addition, the developed information measures can describe accurately the evolution of a SAM over time. Interpreting the SAM matrix as an ergodic chain could show the effect of a shock on the economy after several periods or economic cycles. Under our new framework, finding the power limit of the matrix allows one to check (and confirm) whether the matrix is well-constructed (irreducible and aperiodic), and obtain new optimization functions to balance the SAM matrix. In addition to the theory, we also provide two empirical examples that support our channel concept and help to understand the associated measures.

Keywords: social accounting matrix; entropy; mutual information; information channel; markov chain

Mutual Information



Mutual Information

$$I(x_i; y_j) = I(x_i) - I(x_i|y_j)$$

$$I(x_i; y_j) = -\log_b p(x_i) - [-\log_b p(x_i|y_j)]$$

$$I(x_i; y_j) = \log_b \frac{p(x_i|y_j)}{p(x_i)}$$

$$I(y_j; x_i) = I(y_j) - I(y_j|x_i)$$

$$I(y_j; x_i) = -\log_b p(y_j) - [-\log_b p(y_j|x_i)]$$

$$I(y_j; x_i) = \log_b \frac{p(y_j|x_i)}{p(y_j)}$$

$$I(x_i; y_j) = I(y_j; x_i)$$

Average Mutual Information

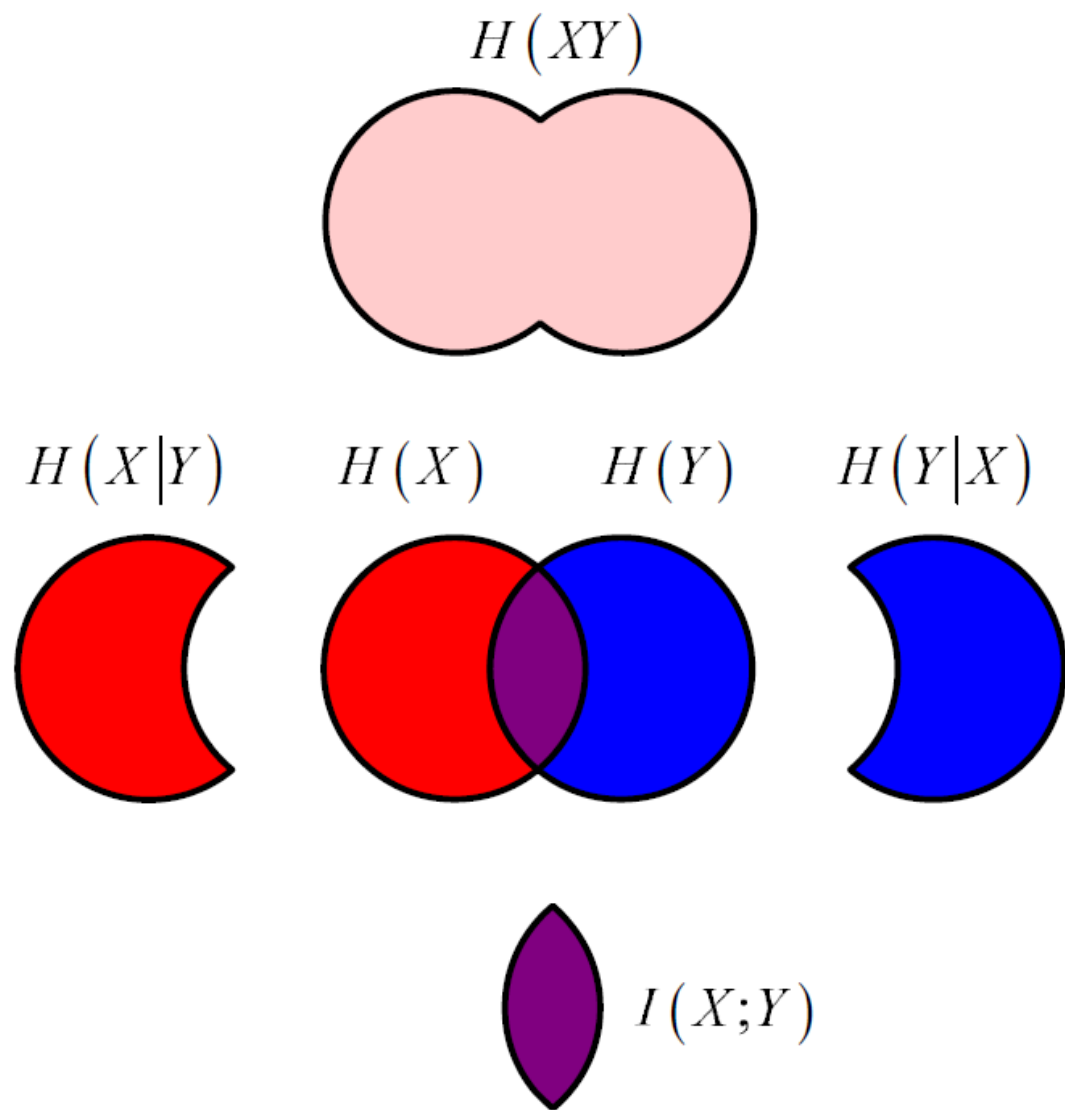
$$I(X; Y) = \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) I(x_i; y_j)$$

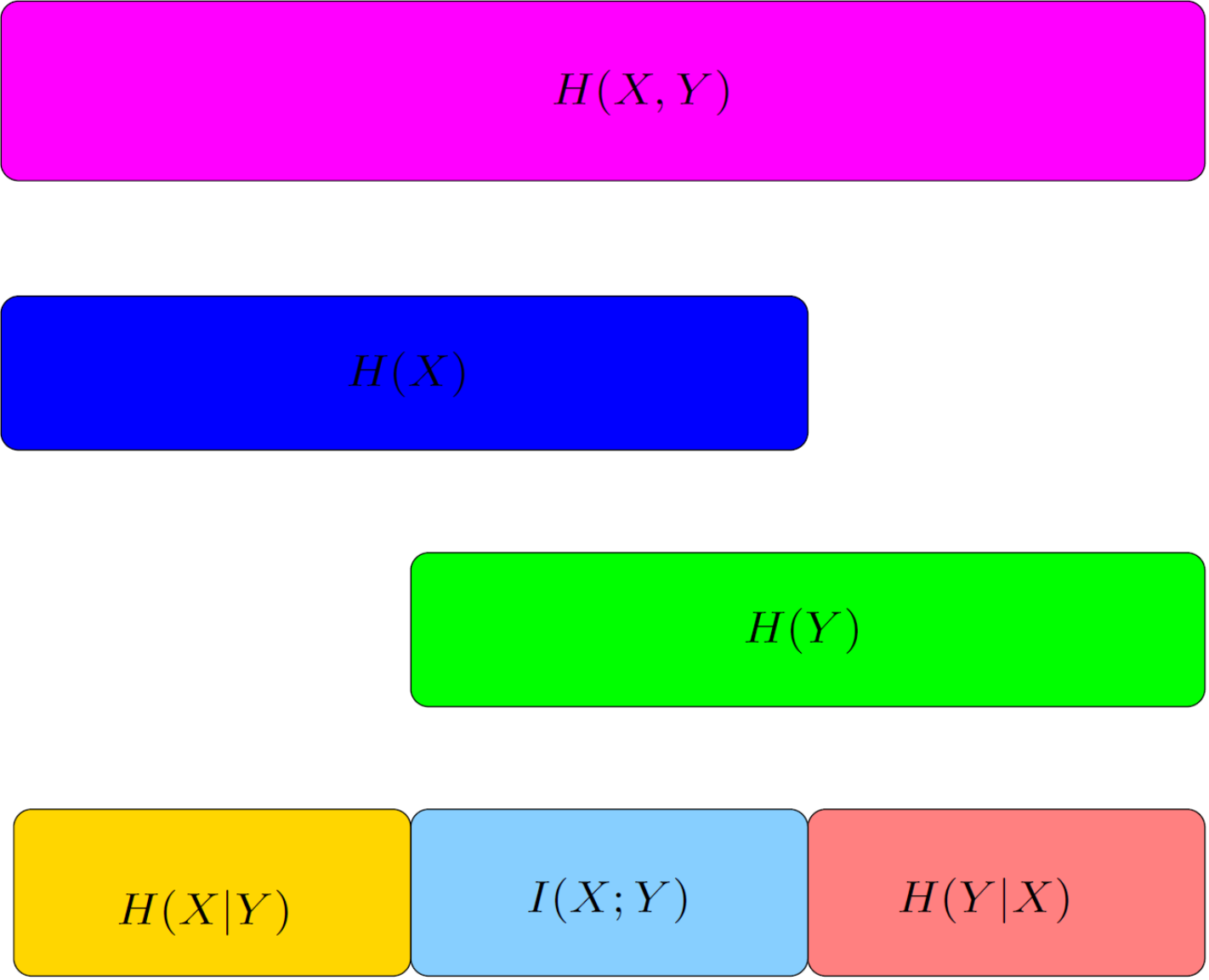
$$I(X; Y) = \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log_b \frac{p(x_i|y_j)}{p(x_i)}$$

Average Mutual Information

$$\begin{aligned} I(X; Y) &= \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log_b \frac{p(x_i|y_j)}{p(x_i)} \\ &= - \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log_b p(x_i) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log_b p(x_i|y_j) \end{aligned}$$

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





A Venn diagram illustrating the relationship between joint and conditional entropies. It consists of two overlapping circles. The left circle is yellow and labeled $H(X|Y)$. The right circle is red and labeled $H(Y|X)$. The intersection of the two circles is shaded light blue and labeled $I(X; Y)$. A large magenta rectangle, labeled $H(X, Y)$, encompasses the entire diagram, representing the total joint entropy.

$$H(X, Y)$$

$$H(X)$$

$$H(Y)$$

$$H(X|Y)$$

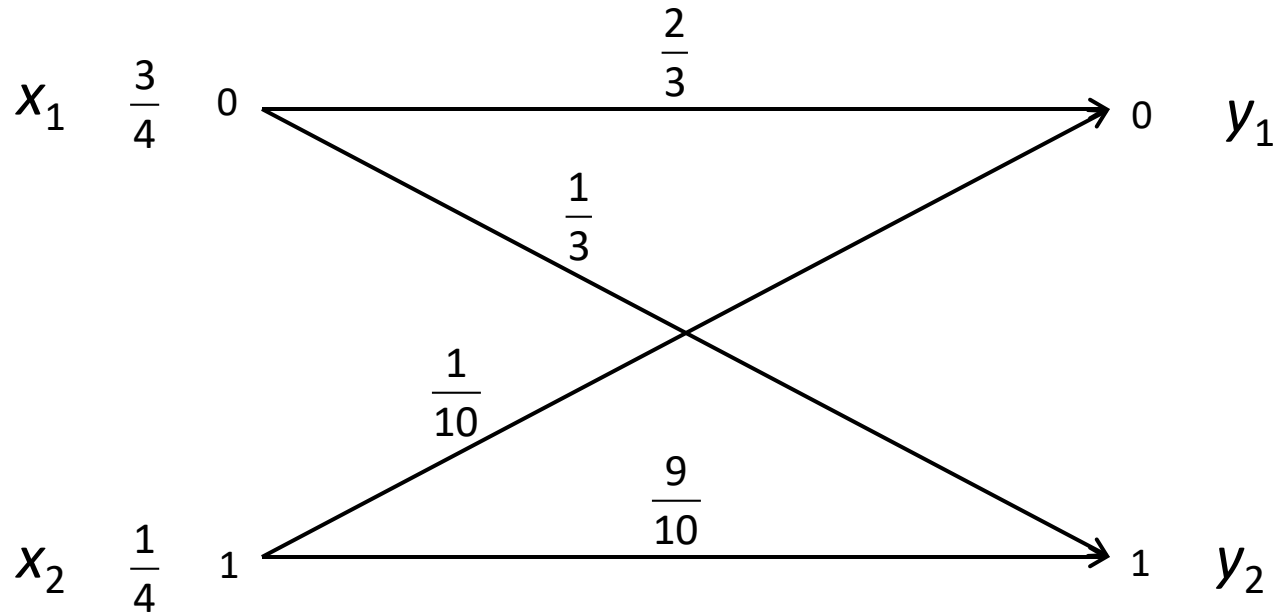
$$I(X; Y)$$

$$H(Y|X)$$

Four Vectors Example

- $[0,0,0], [0,1,0], [1,0,0], [1,0,1]$ (equiprobable)
- $p(y=0) = .75, p(y=1) = .25 \rightarrow H(Y) = .811 \text{ bit}$
- $I(X;Y) = H(Y) - H(Y|X)$
 $= .811 - .50 = .311 \text{ bit}$
- $p(x=0) = p(x=1) = .50 \rightarrow H(X) = 1.0 \text{ bit}$
- $I(X;Y) = H(X) - H(X|Y)$
 $= 1.0 - .689 = .311 \text{ bit}$
- $H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y) = 1.5 \text{ bits}$

Non-symmetric Binary Channel



channel matrix $\begin{bmatrix} p(y_1 | x_1) & p(y_2 | x_1) \\ p(y_1 | x_2) & p(y_2 | x_2) \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$

Non-symmetric Channel Example

$$I(X;Y) = .192 \text{ bit}$$

$$H(X) = .811 \text{ bit}$$

- $H(X|Y) = H(X) - I(X;Y) = .619 \text{ bit}$

$$H(Y) = .998 \text{ bit}$$

- $H(Y|X) = H(Y) - I(X;Y) = .806 \text{ bit}$

- $H(XY) = H(X) + H(Y|X)$
 $= H(Y) + H(X|Y) = 1.617 \text{ bits}$

Mutual Information for the BSC

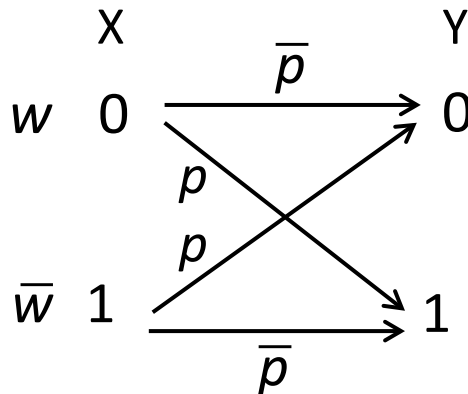


crossover probability p

$$\bar{p} = 1 - p$$

channel matrix

$$\begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$



$$p(x = 0) = w$$

$$p(x = 1) = 1 - w = \bar{w}$$

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

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

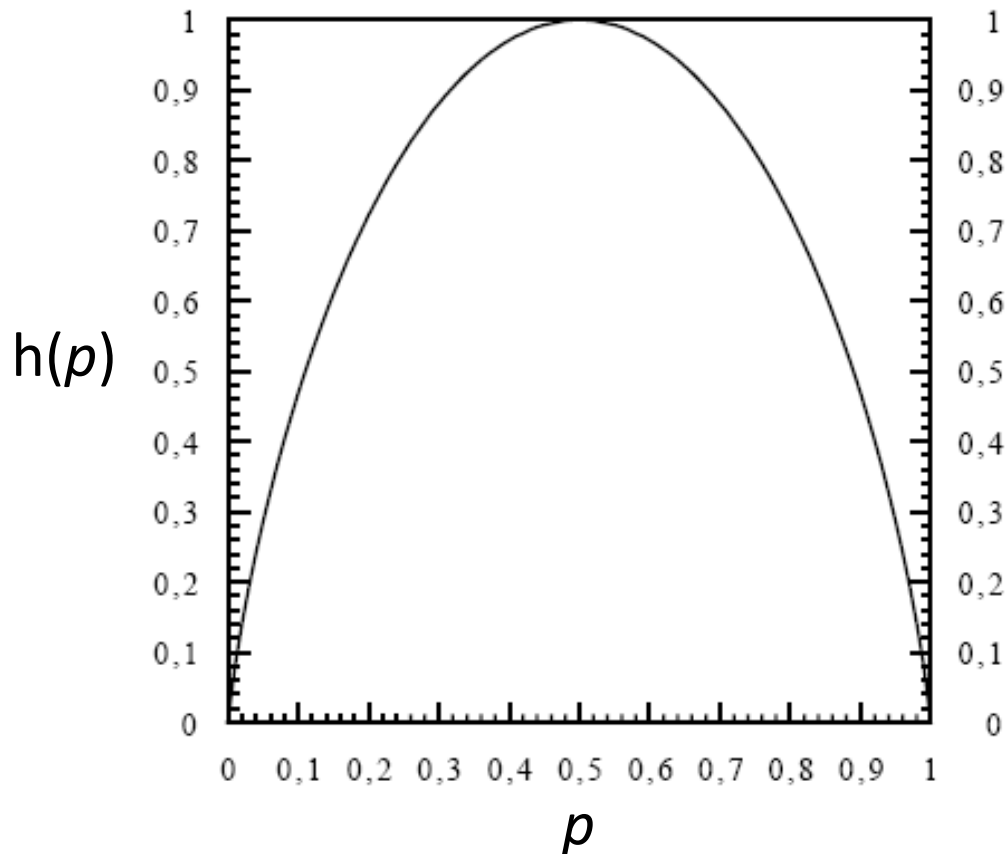
Mutual Information for the BSC

$$\begin{aligned} H(Y|X) &= - \sum_i \sum_j p(x_i, y_j) \log p(y_j|x_i) \\ &= - \sum_i \sum_j p(x_i) p(y_j|x_i) \log p(y_j|x_i) \\ &= - \sum_i p(x_i) \sum_j p(y_j|x_i) \log p(y_j|x_i) \\ &= - \sum_i p(x_i) [p \log p + \bar{p} \log \bar{p}] \\ &= - [p \log p + \bar{p} \log \bar{p}] \end{aligned}$$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) + [p \log p + \bar{p} \log \bar{p}] \\ &= H(Y) - h(p) \end{aligned}$$

Binary Entropy Function

$$h(p) = - [p \log_2 p + (1-p) \log_2 (1-p)] \quad 0 \leq p \leq 1$$



$$p(y = 0) = w\bar{p} + \bar{w}p$$

$$p(y = 1) = wp + \bar{w}\bar{p}$$

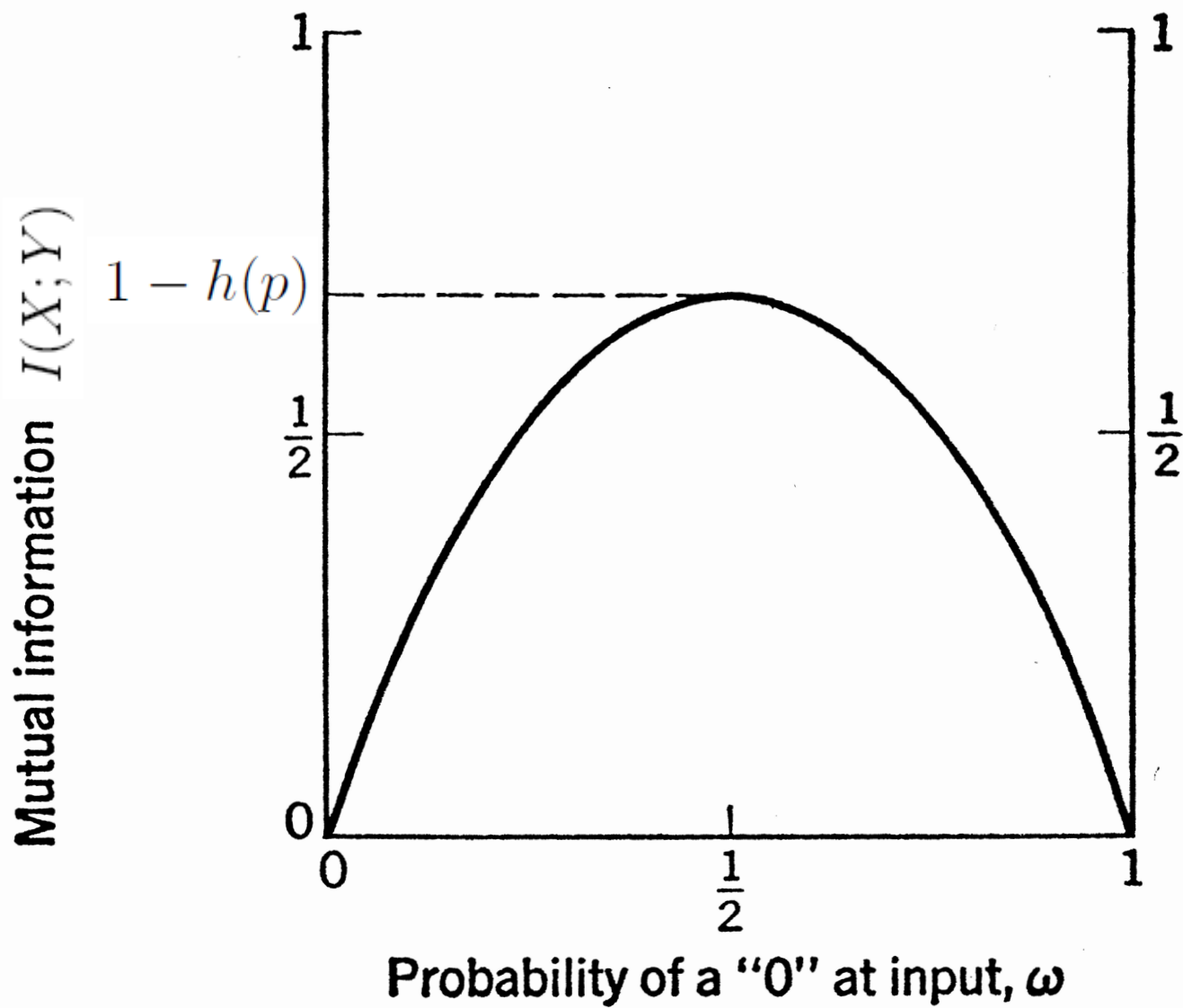
$$H(Y) = -[(w\bar{p} + \bar{w}p) \log(w\bar{p} + \bar{w}p) + (wp + \bar{w}\bar{p}) \log(wp + \bar{w}\bar{p})]$$

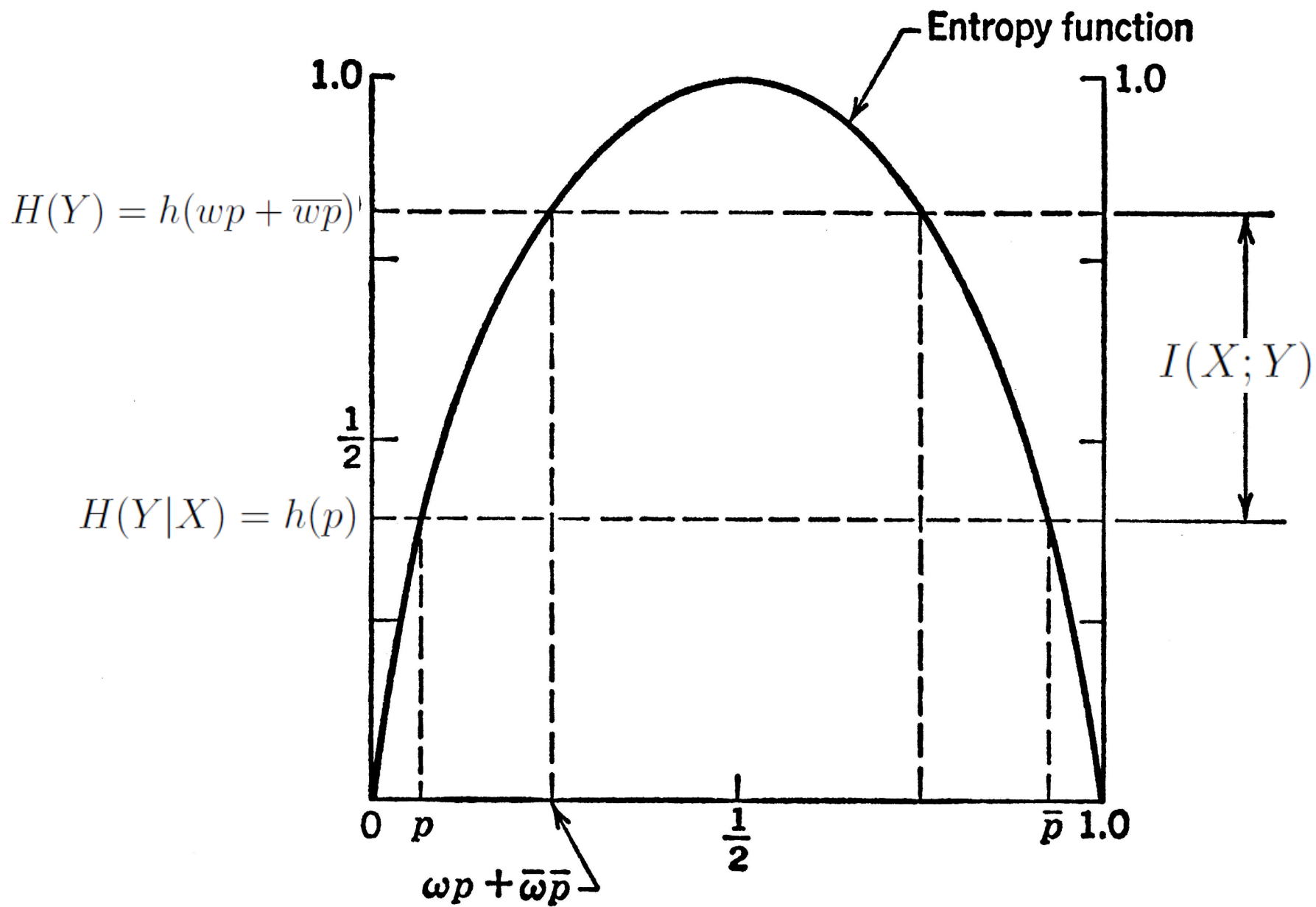
For $w = 1/2$

$$\begin{aligned} I(X; Y) &= h[1/2(p) + 1/2(1 - p)] - h(p) \\ &= h(1/2) - h(p) \\ &= 1 - h(p) \end{aligned}$$

For $w = 0$ or 1

$$\begin{aligned} I(X; Y) &= h(p) - h(p) \\ &= 0 \end{aligned}$$





Midterm Test

- Wednesday, October 28, 2026
- During class time (10:30 – 11:20)
- Counts for 25% of the final mark
- Aids allowed
 - One page of notes on 8.5" × 11.5" paper (both sides)
 - Calculator
- Cellphones, tablets, laptops, or any other electronic devices are NOT ALLOWED

