ELEC 519A
Selected Topics in Digital Communications: Information Theory

Joint Entropy, Equivocation and Mutual Information

Joint Entropy

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

Conditional Entropy

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

Chain Rule

$$\boxed{H(XY) = H(X) + H(Y|X)}$$

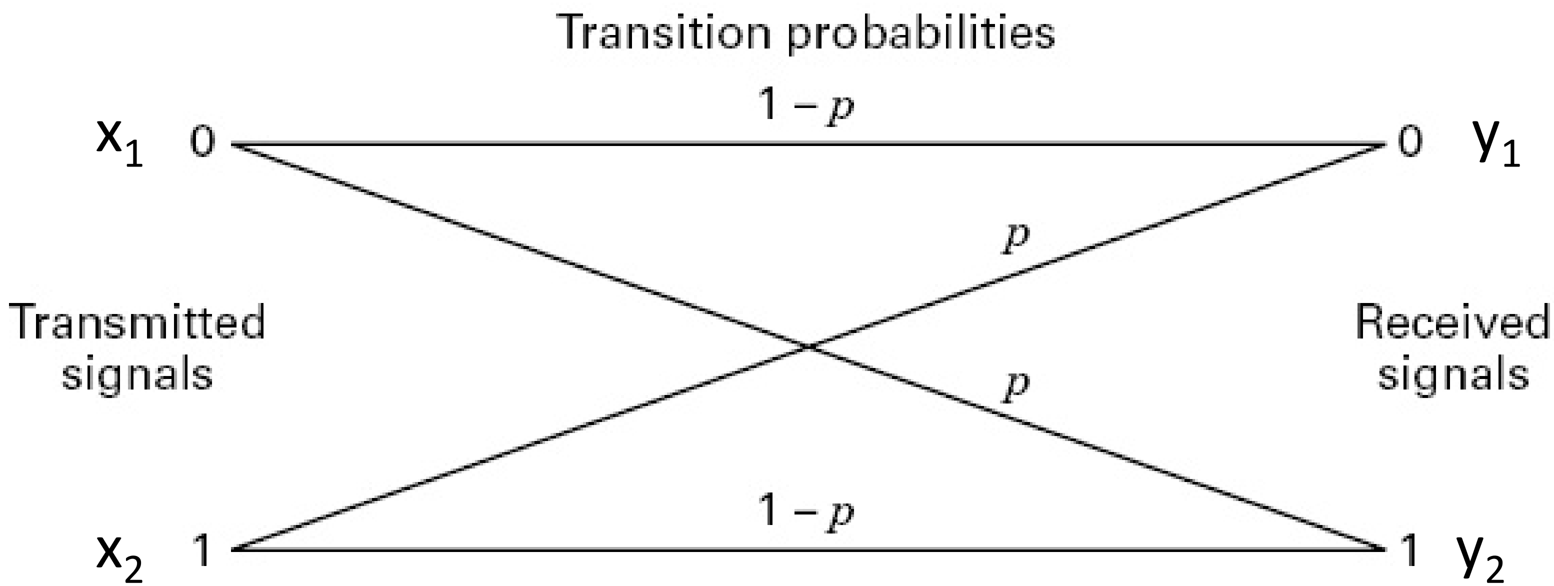
$$\boxed{H(X_1, \ldots, X_N) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots + H(X_N|X_1, \ldots, X_{N-1})}$$

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

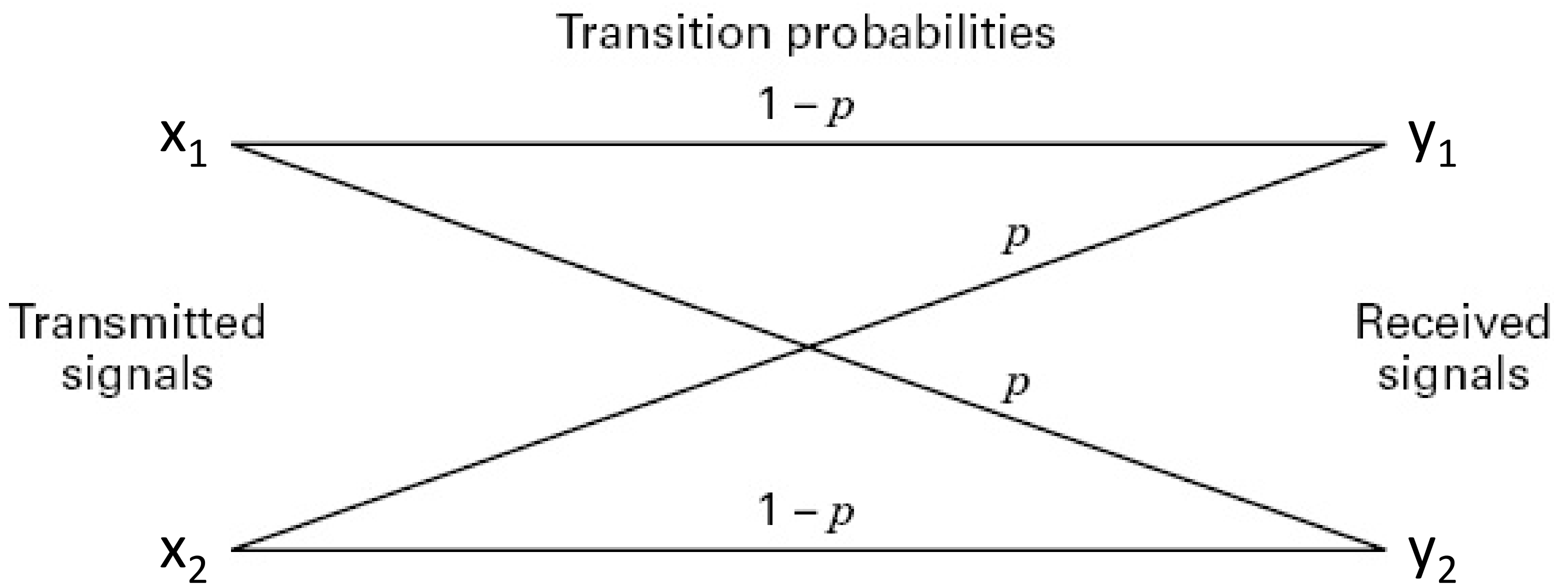
Information Channels

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

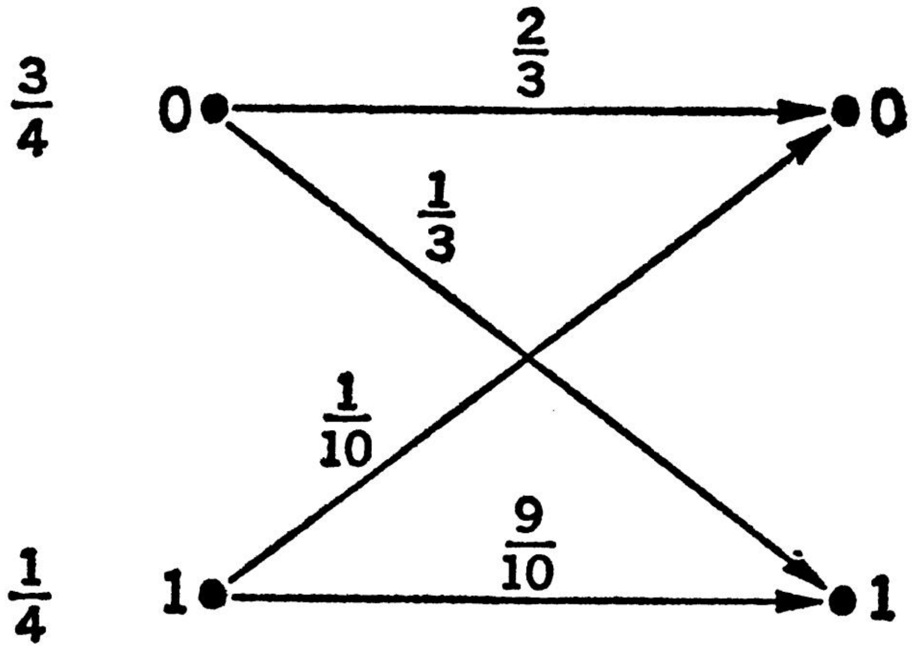
Binary Symmetric Channel



Binary Symmetric Channel



Non-Symmetric Binary Channel



Mutual Information

$$
\begin{aligned}
I(x_i; y_j) &= I(x_i) - I(x_i|y_j) \\
I(x_i; y_j) &= -\log_b p(x_i) - [-\log_b p(x_i|y_j)]
\end{aligned}
$$

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

$$
\begin{aligned}
I(y_j; x_i) &= I(y_j) - I(y_j|x_i) \\
I(y_j; x_i) &= -\log_b p(y_j) - [-\log_b p(y_j|x_i)]
\end{aligned}
$$

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

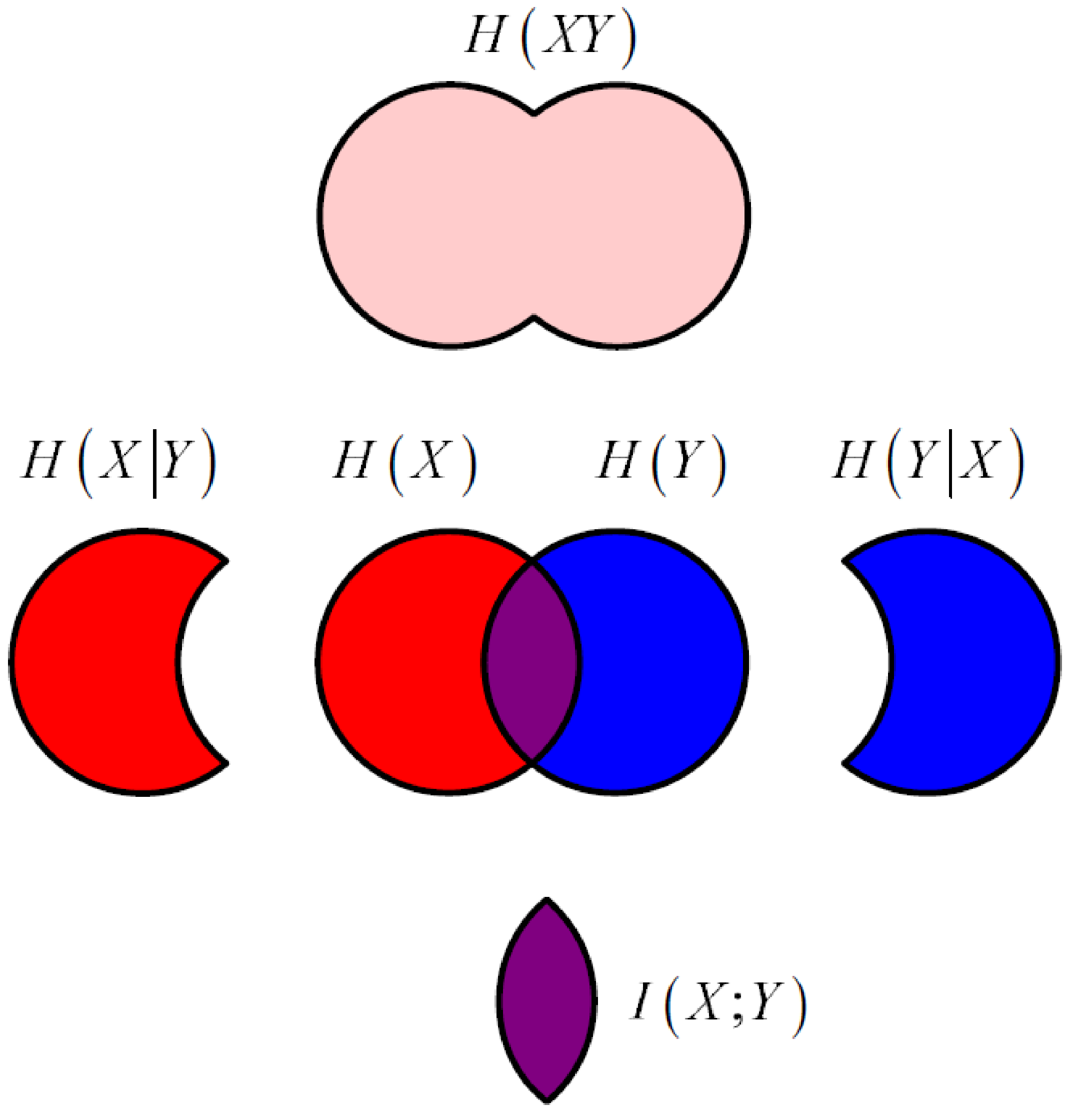
$$\boxed{I(x_i; y_j) = I(y_j; x_i)}$$

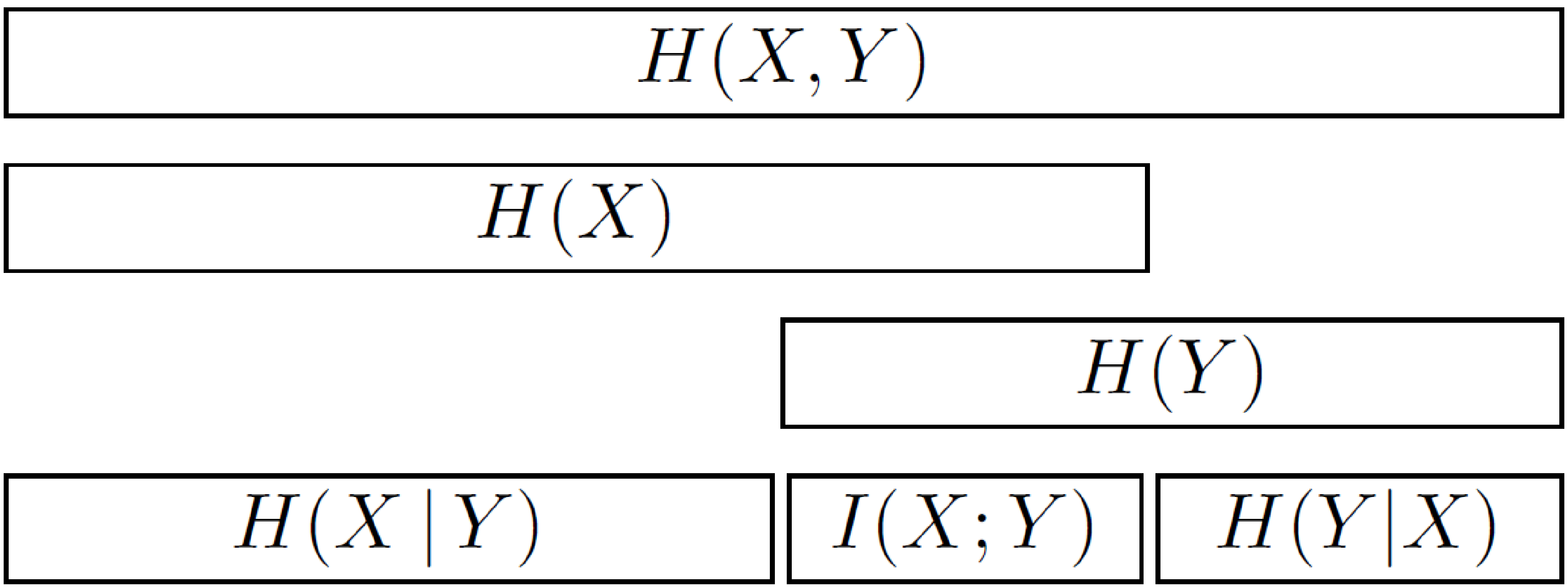
Average Mutual Information

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

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

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


$H(XY)$
$H(X|Y)$
$H(X)$
$H(Y)$
$H(Y|X)$
$I(X;Y)$

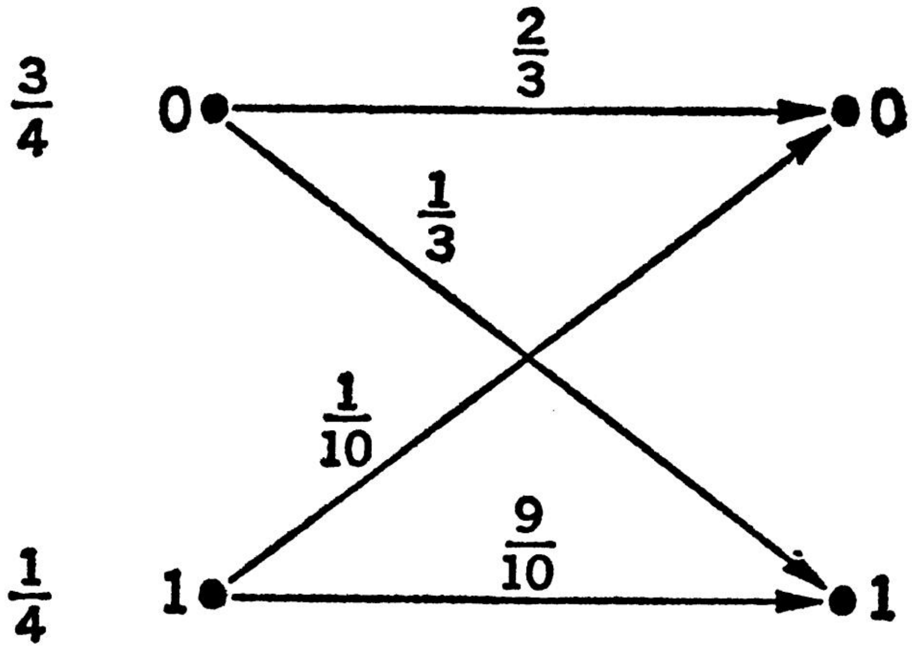

$H(X,Y)$
$H(X)$
$H(Y)$
$H(X \mid Y)$
$I(X;Y)$
$H(Y|X)$

Example

- [0,0,0], [0,1,0], [1,0,0], [1,0,1]
- $P_Y(1) = \frac{1}{4} \rightarrow H(Y) = .811$ bits
- $I(X;Y) = H(Y) - H(Y|X)$

  $= .811 - .500 = .311$ bits

Non-Symmetric Binary Channel



Non-Symmetric Channel Example

$H(X) \quad = .811$

$H(X|Y) \quad = .619$

- $H(X) - H(X|Y) = I(X;Y) = .192$

$H(Y) \quad = .998$

$H(Y|X) \quad = .806$

- $H(Y) - H(Y|X) = I(Y;X) = .192$

- $H(XY) = H(X) + H(Y|X) = H(Y)+H(X|Y) = 1.617$

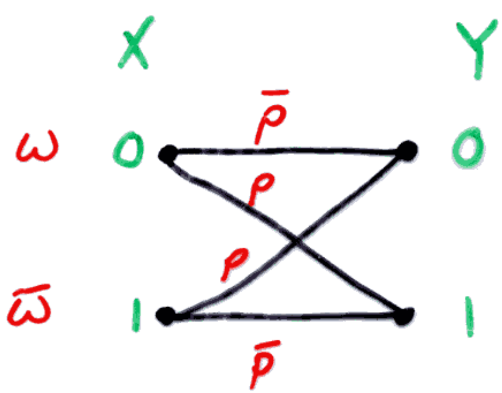
Mutual Information for a BSC

$X \to \boxed{BSC} \to Y$ crossover probability p

$\bar{p} = 1 - p$

channel matrix

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



$p(X=0) = \omega$

$p(X=1) = 1 - \omega = \bar{\omega}$

Mutual Information for a BSC

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

$$I(X;Y) = H(Y) + [\, p \log p + \overline{p} \log \overline{p}\,]$$

$$
\begin{aligned}
p(y=0) &= w\overline{p} + \overline{w}p \\
p(y=1) &= wp + \overline{wp}
\end{aligned}
$$

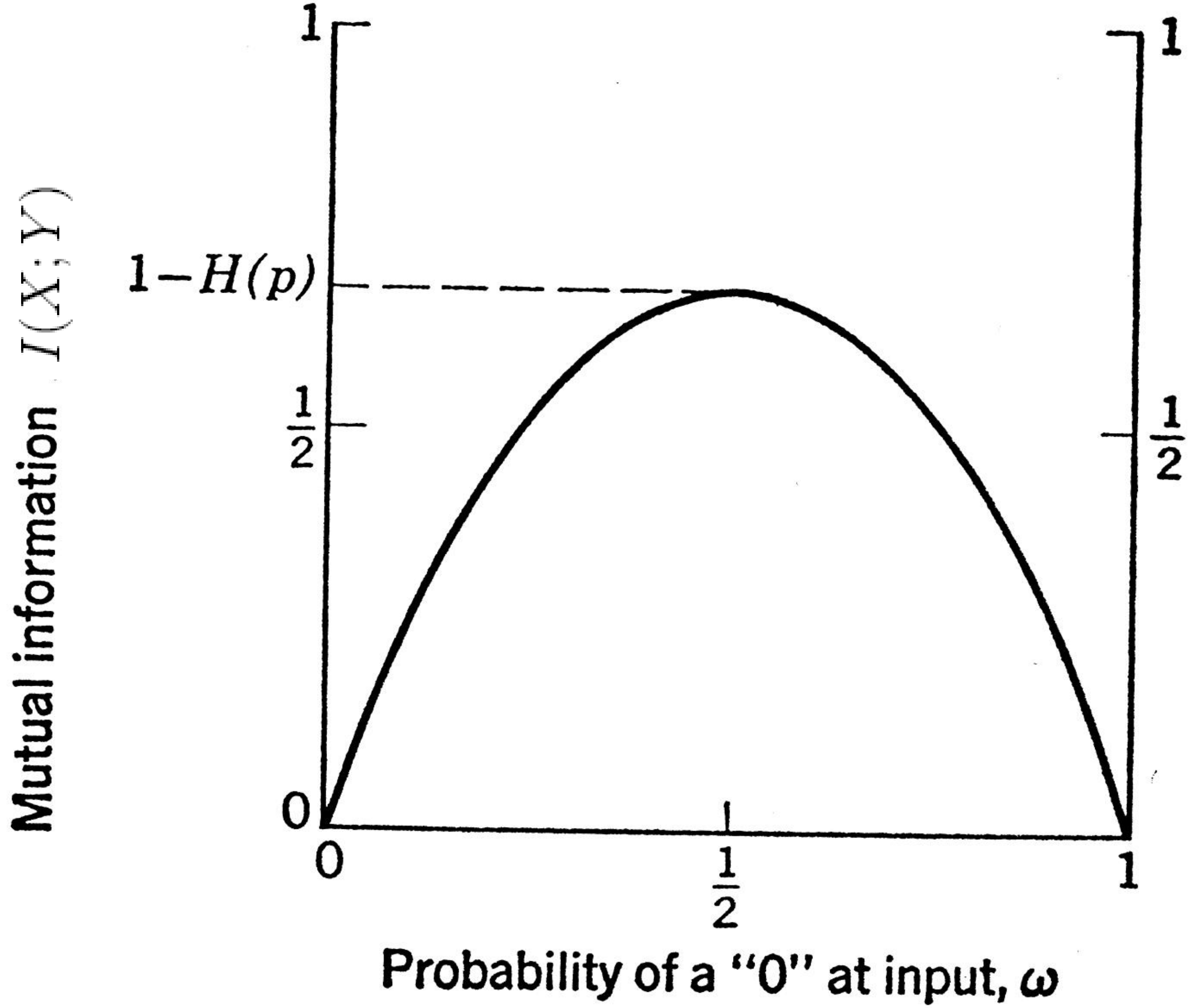
$$
\begin{aligned}
H(Y) &= -[\,(w\overline{p} + \overline{w}p)\log(w\overline{p} + \overline{w}p) + (wp + \overline{wp})\log(wp + \overline{wp})\,] \\
&= H(wp + \overline{wp}) \\
I(X;Y) &= H(wp + \overline{wp}) - H(p)
\end{aligned}
$$

For $w = 1/2$

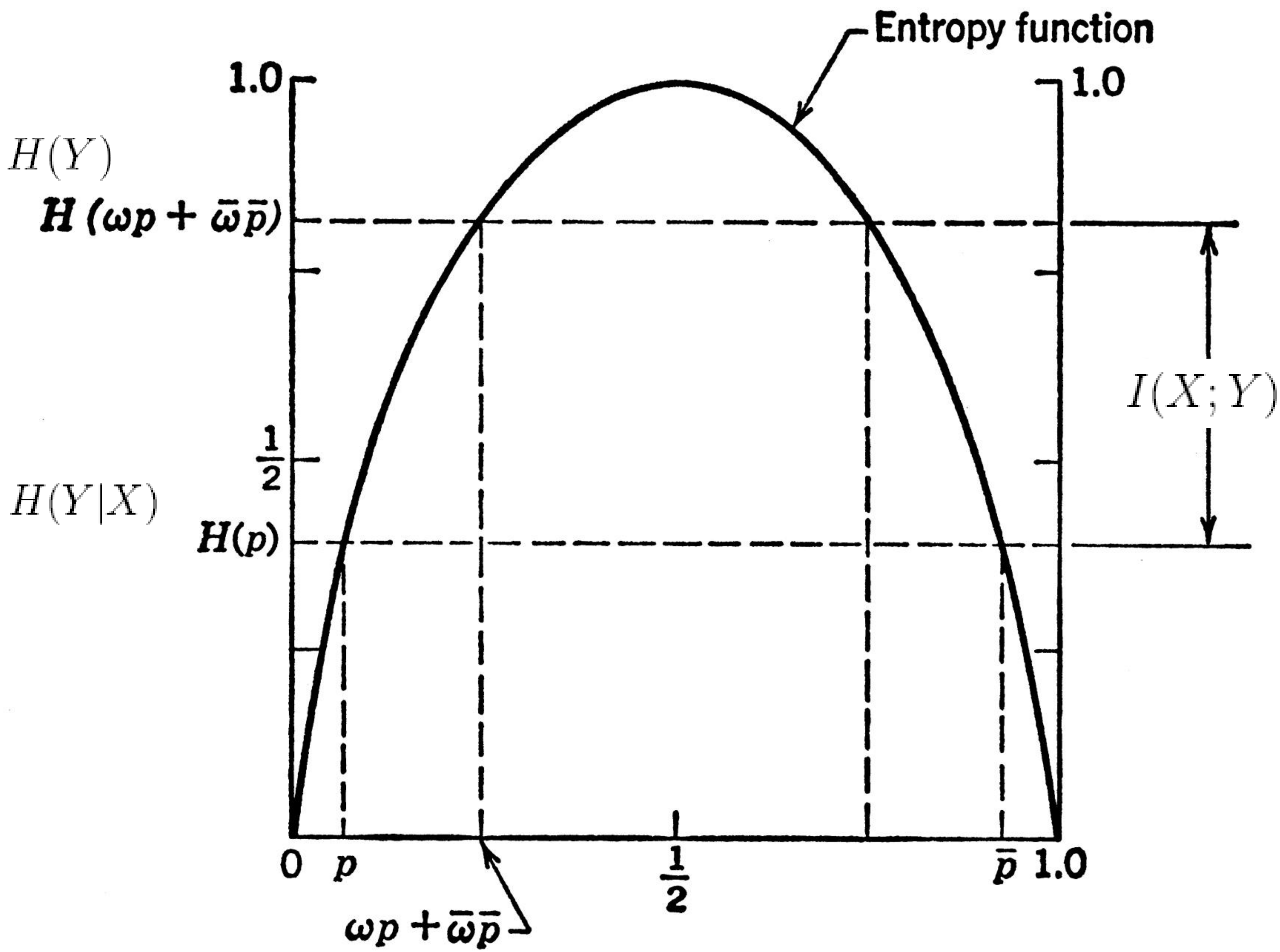
$$
\begin{aligned}
I(X:Y) &= H[1/2(p) + 1/2(1-p)] - H(p) \\
&= H(1/2) - H(p) \\
&= 1 - H(p)
\end{aligned}
$$

For $w = 0$ or 1

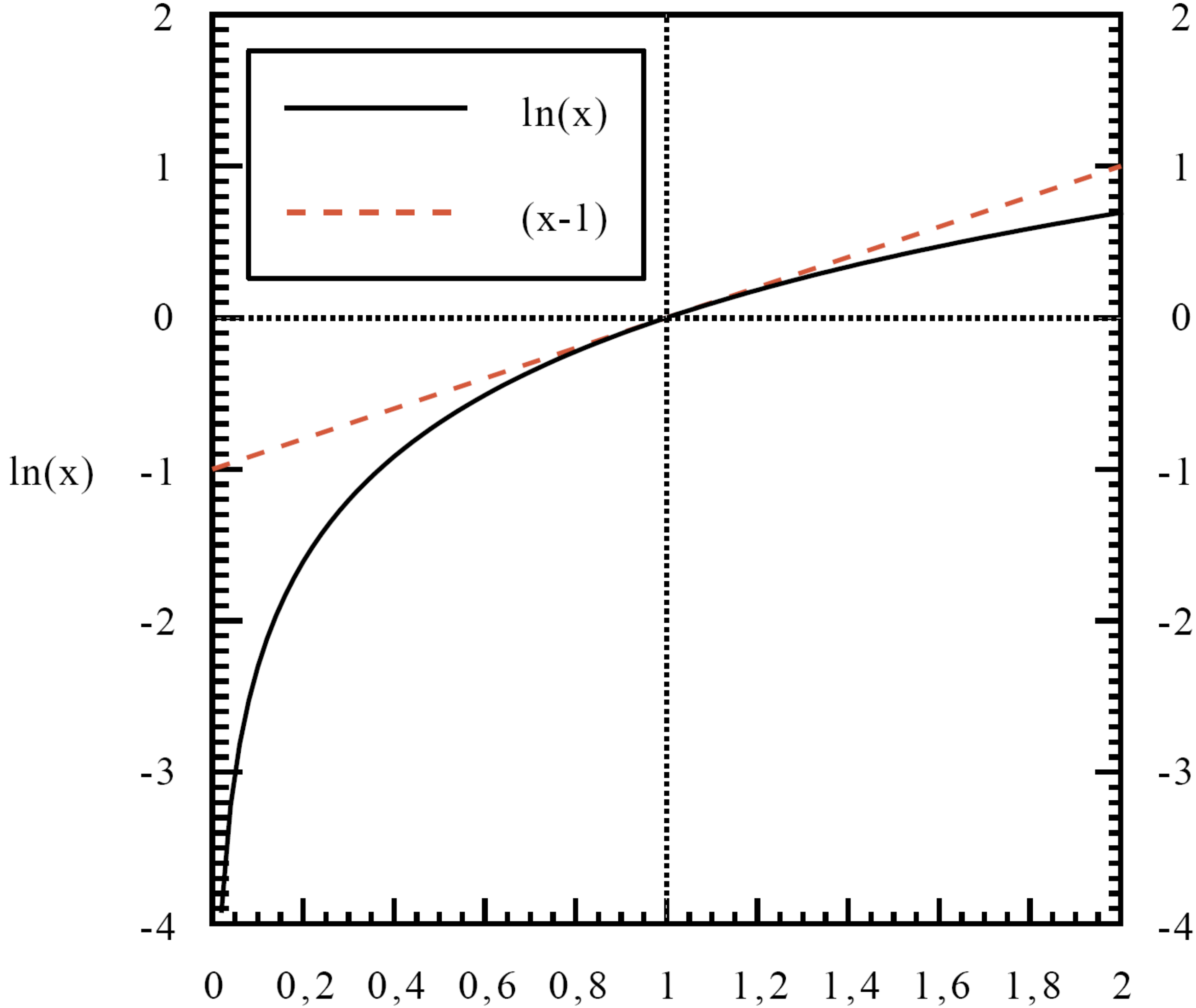
$$
\begin{aligned}
I(X:Y) &= H(p) - H(p) \\
&= 0
\end{aligned}
$$



Mutual information $I(X;Y)$
1
$1-H(p)$
$\frac{1}{2}$
0
0
$\frac{1}{2}$
1
1
$\frac{1}{2}$
Probability of a "0" at input, ω



Entropy function
1.0
1.0
$H(Y)$
$H(\omega p + \bar{\omega}\bar{p})$
$I(X;Y)$
$\frac{1}{2}$
$H(Y|X)$
$H(p)$
0
p
$\frac{1}{2}$
$\bar{p}$
1.0
$\omega p + \bar{\omega}\bar{p}$


ln(x)
(x-1)
ln(x)
2
1
0
-1
-2
-3
-4
0
0,2
0,4
0,6
0,8
1
1,2
1,4
1,6
1,8
2

Conditional Mutual Information

$$\boxed{I(x_i; y_j | z_k) \equiv \log_b \frac{p(x_i | y_j, z_k)}{p(x_i | z_k)}}$$

Conditional Mutual Information

$$\boxed{I(x_i; y_j | z_k) = I(x_i | z_k) - I(x_i | y_j, z_k)}$$

Conditional Mutual Information

I(X;Y) = H(X) – H(X|Y)

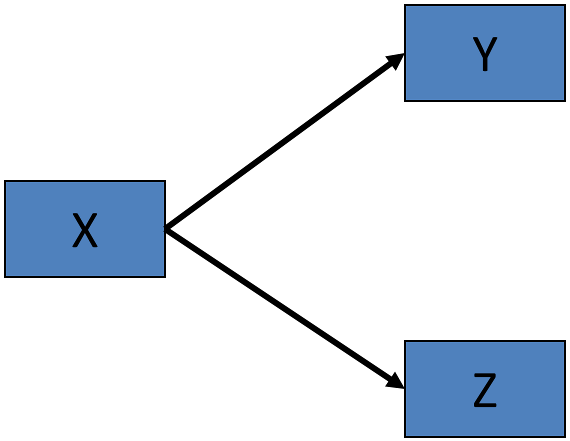
$H(X|Z)$: average uncertainty remaining in X after the observation in Z

$H(X|YZ)$: average uncertainty remaining in X after the observation in both Z and Y

$I(X;Y|Z)$: average amount of uncertainty in X *resolved* by the observation in Y

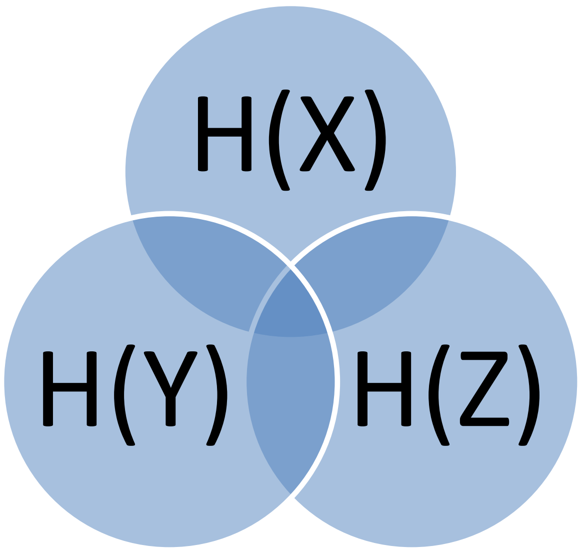
I(X;Y|Z) = H(X|Z) – H(X|YZ)

Joint Mutual Information



$$I(X;YZ) = I(X;Y) + I(X;Z|Y)$$
$$I(X;YZ) = I(X;Z) + I(X;Y|Z)$$
$$I(X;YZ) \leq I(X;Y) + I(X;Z)$$

H(XYZ)



Mutual Information for Three Alphabets

- For two alphabets

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

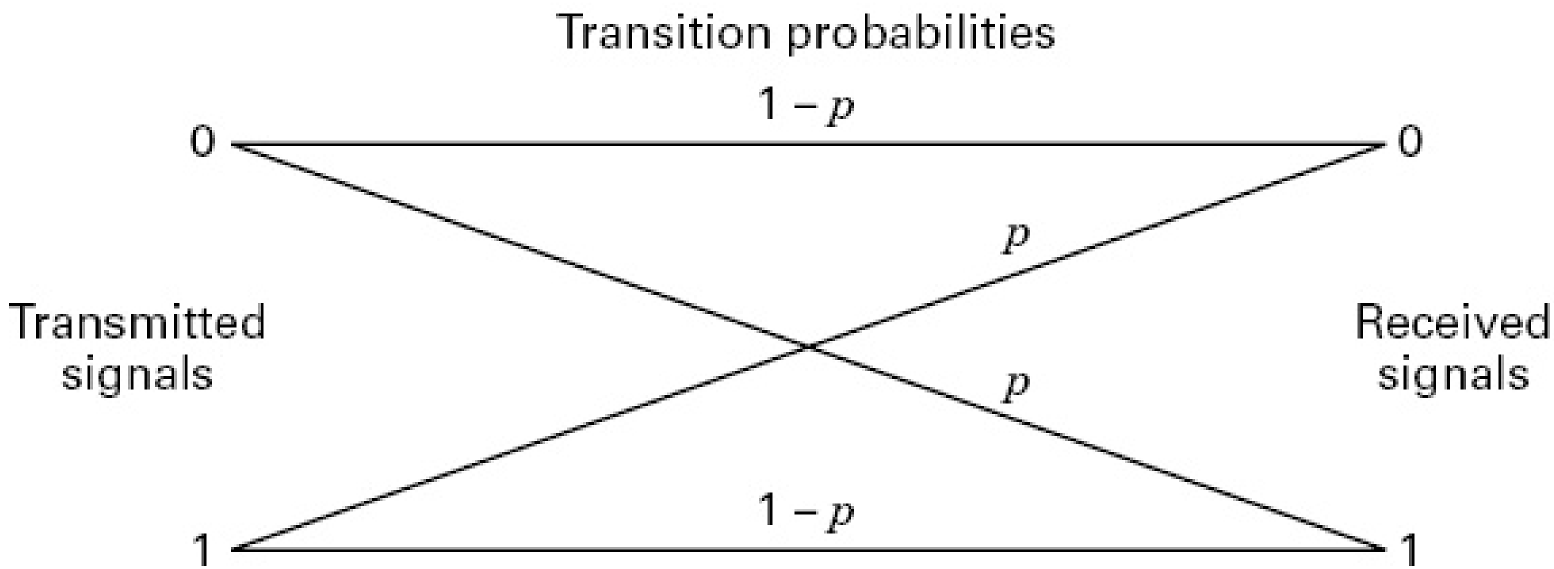
- For three alphabets

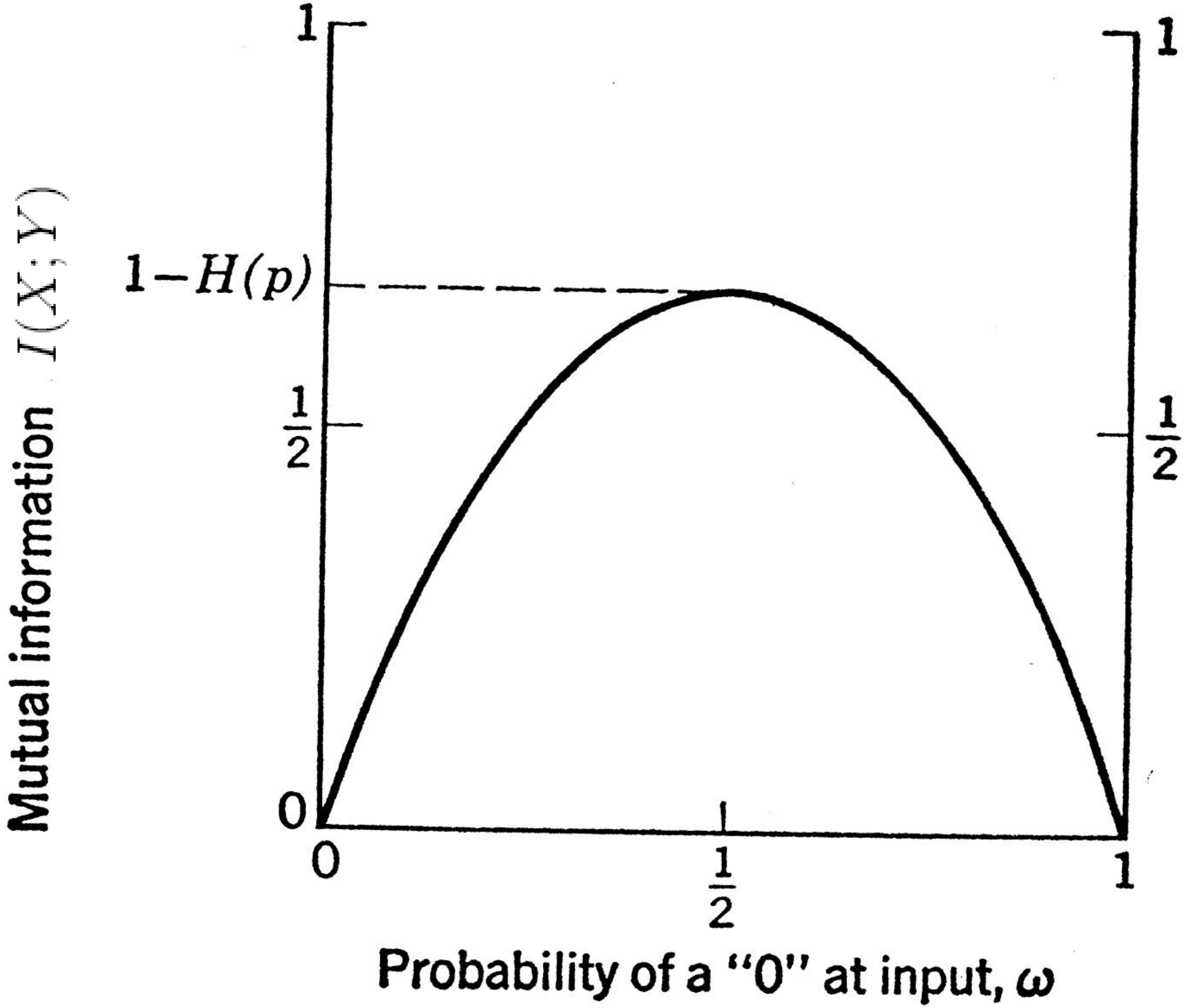
  $$I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$$

Probabilities for Three RVs

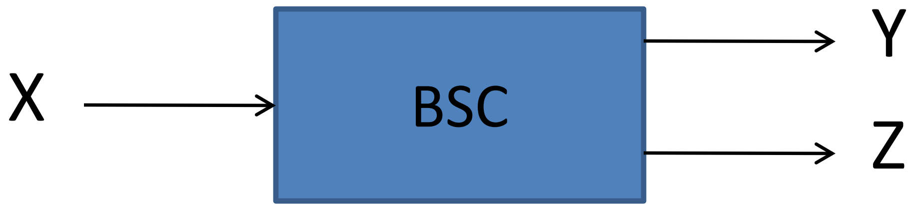
x_i	y_j	z_k	$p(x_i, y_j, z_k)$	$p(x_i \mid y_j, z_k)$	$p(x_i \mid z_k)$	$p(x_i, y_j)$	$p(x_i)$
0	0	0	1/4	1	1/2	1/4	1/2
0	0	1	0	0	1/2	1/4	1/2
0	1	0	0	0	1/2	1/4	1/2
0	1	1	1/4	1	1/2	1/4	1/2
1	0	0	0	0	1/2	1/4	1/2
1	0	1	1/4	1	1/2	1/4	1/2
1	1	0	1/4	1	1/2	1/4	1/2
1	1	1	0	0	1/2	1/4	1/2

Binary Symmetric Channel



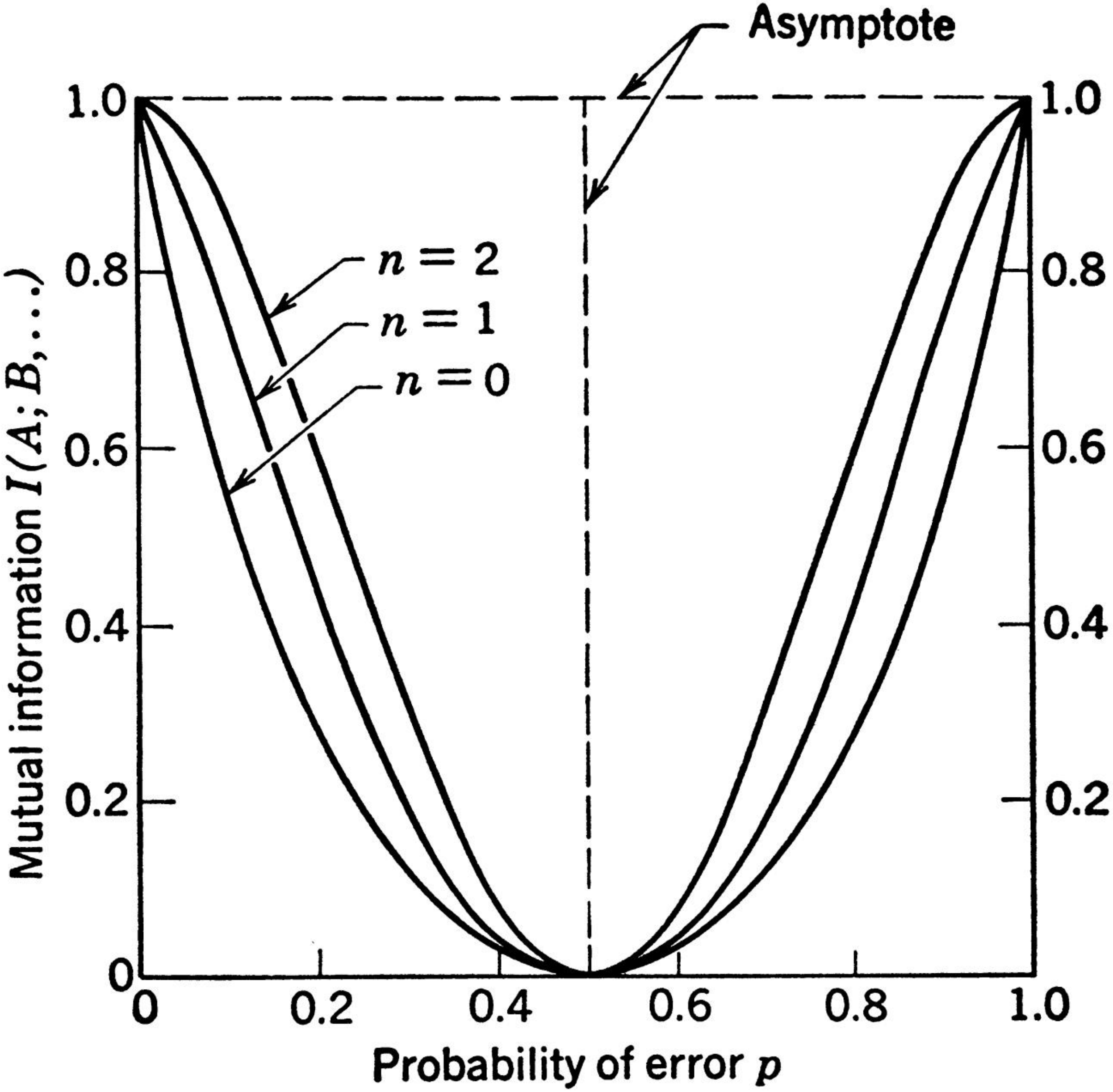

1
1
$1-H(p)$
$\frac{1}{2}$
$\frac{1}{2}$
Mutual information $I(X;Y)$
0
0
$\frac{1}{2}$
1
Probability of a "0" at input, ω

Additivity of Mutual Information



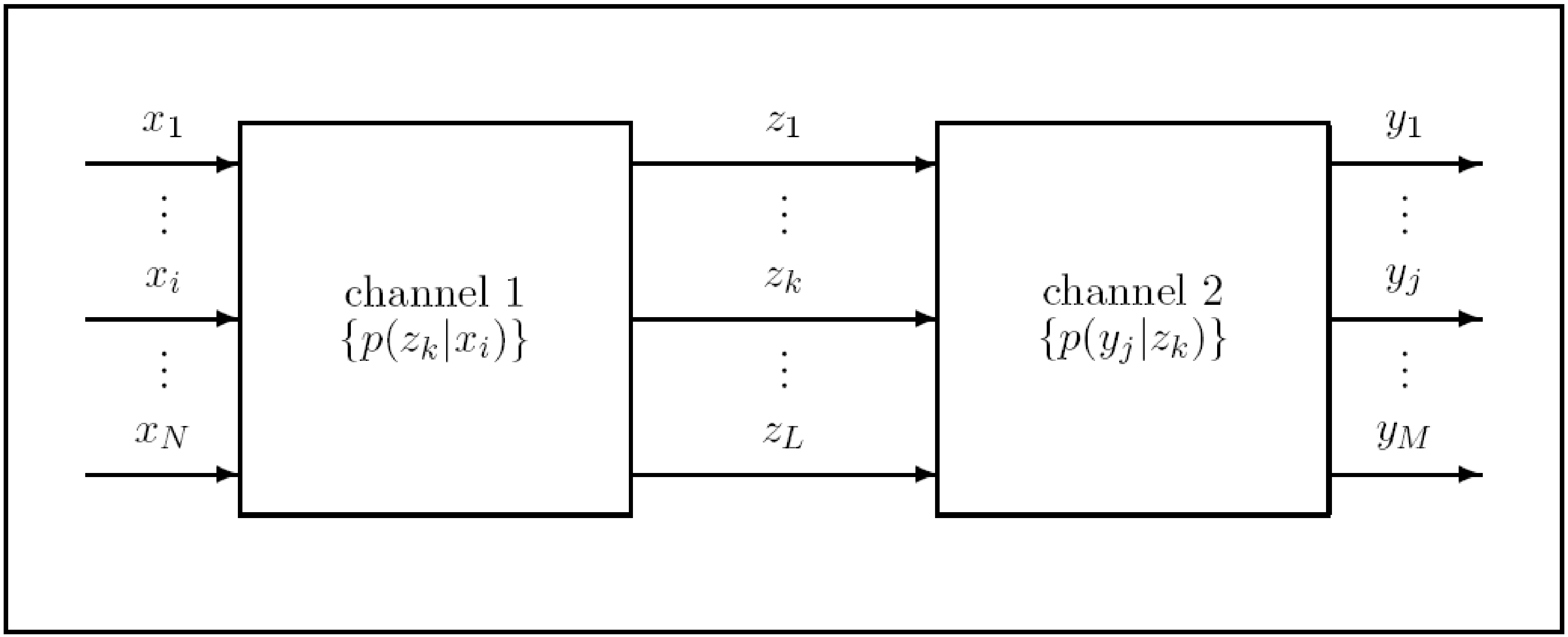
Probabilities for a Repetitive BSC

x_i	y_j	z_k	$p(x_i)$	$p(x_i, y_j, z_k)$	$p(y_j, z_k)$
0	0	0	$1/2$	$1/2(\overline{p}^2)$	$1/2(p^2 + \overline{p}^2)$
0	0	1	$1/2$	$1/2(p\overline{p})$	$p\overline{p}$
0	1	0	$1/2$	$1/2(p\overline{p})$	$p\overline{p}$
0	1	1	$1/2$	$1/2(p^2)$	$1/2(p^2 + \overline{p}^2)$
1	0	0	$1/2$	$1/2(p^2)$	$1/2(p^2 + \overline{p}^2)$
1	0	1	$1/2$	$1/2(p\overline{p})$	$p\overline{p}$
1	1	0	$1/2$	$1/2(p\overline{p})$	$p\overline{p}$
1	1	1	$1/2$	$1/2(\overline{p}^2)$	$1/2(p^2 + \overline{p}^2)$



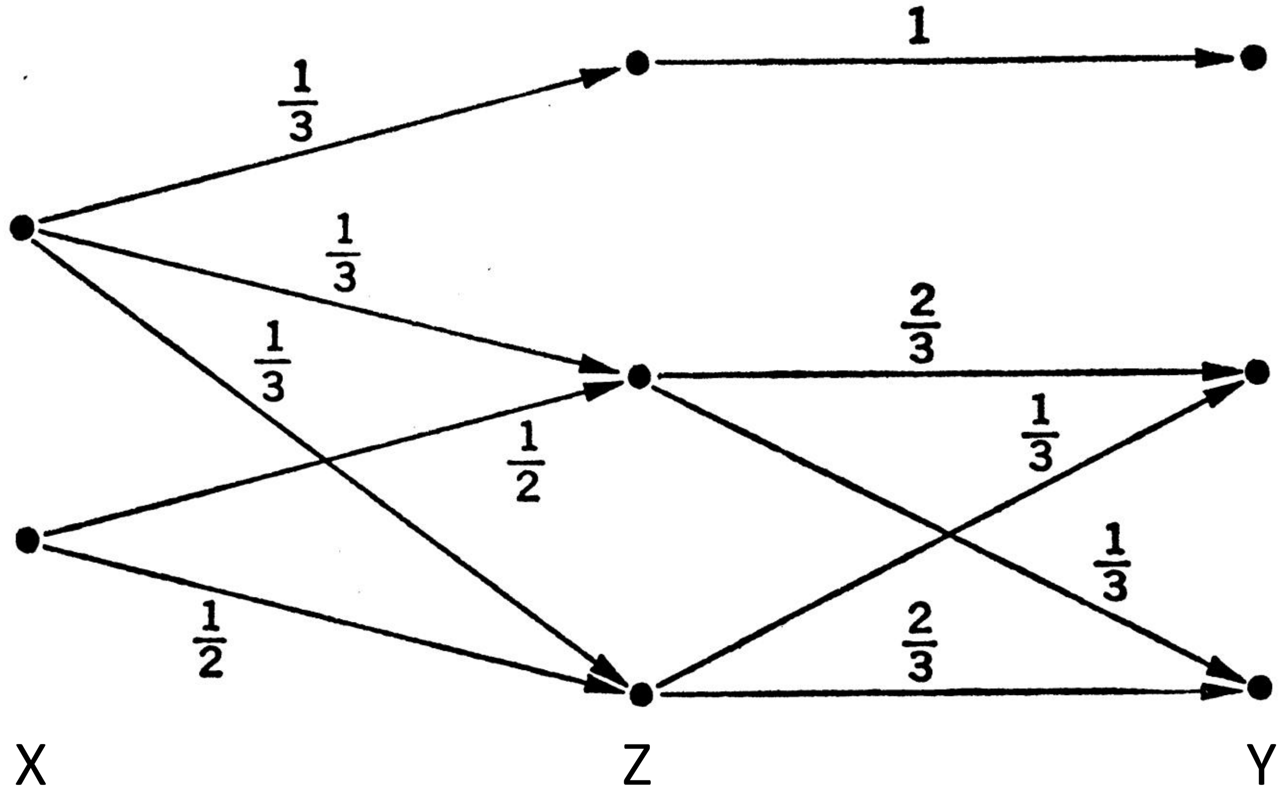
Asymptote
1.0
0.8
0.6
0.4
0.2
0
$n = 2$
$n = 1$
$n = 0$
Mutual information $I(A; B, \ldots)$
0
0.2
0.4
0.6
0.8
1.0
Probability of error p

Two Cascaded Channels

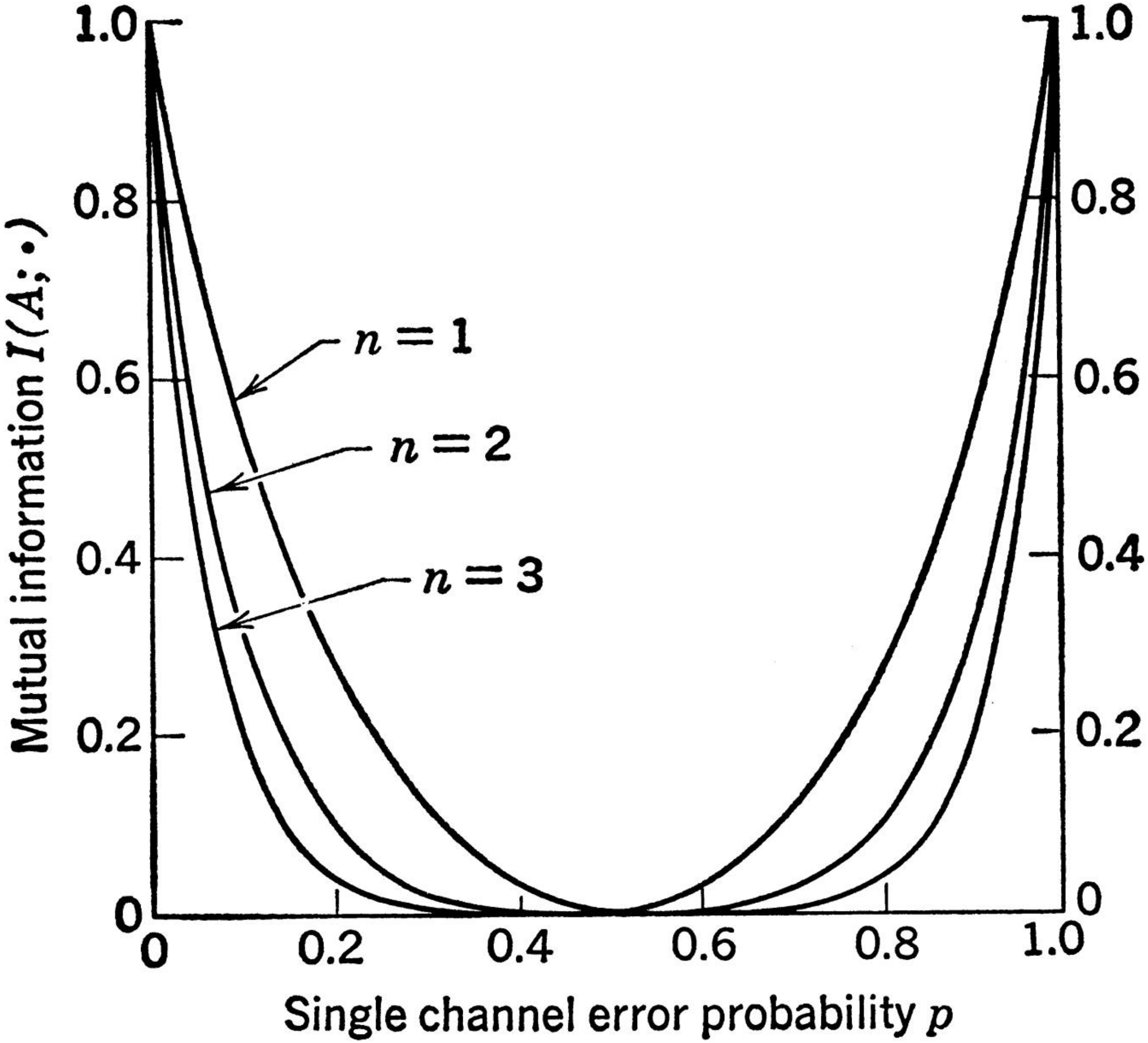


Data Processing Theorem

- The mutual information between the input and output can never exceed the mutual information between the input and an intermediate point
- The mutual information between the input and output can never exceed the mutual information between the output and an intermediate point
- **Data processing cannot increase the amount of information in the data**



1
1/3
1/3
1/3
2/3
1/3
1/2
1/3
1/2
2/3
X
Z
Y



1.0
0.8
0.6
0.4
0.2
0
1.0
0.8
0.6
0.4
0.2
0
$n = 1$
$n = 2$
$n = 3$
Mutual information $I(A; \bullet)$
0
0.2
0.4
0.6
0.8
1.0
Single channel error probability p

A Mathematical Theory of Communications, BSTJ July, 1948

``The fundamental problem of communication is that of reproducing at one point exactly or approximately a message selected at another point. … If the channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with certainty by any operation on the received signal.''

A Mathematical Theory of Communications, BSTJ July, 1948

沟通的根本问题是在一个点上的音响完全或约在另一点选择的消息。如果通道是嘈杂的，它不是一般的可能重建原始邮件或任何操作上接收到的信号传输信号与确定性

``The fundamental problem of communications is a point in the sound completely or choose another point about the news. If the channel is noisy, it is not generally possible to reconstruct the original message or any operation on the received transmission signal and the signal uncertainty.’’

A Mathematical Theory of Communications, BSTJ July, 1948

المشكلة الأساسية للاتصالات نقطة في الصوت كليا أو اختيار نقطة أخرى حول هذا النبأ. وإذا كانت القناة الصاخبة، فإنه ليس منالممكن عموما لإعادة الرسالة الأصلية أو أي عملية على نقل الإشارات التي وردت إشارة عدم اليقين

``The fundamental problem of communication points in the sound completely or choose another point on the news. If the channel is noisy, it is not generally possible to restore the original message or any process on the transfer of signals received signal uncertainty.''

A Mathematical Theory of Communications, BSTJ July, 1948

Das grundlegende Problem der Kommunikation Punkte in den Sound komplett oder wählen Sie einen anderen Punkt auf die Nachricht. Wenn der Kanal ist laut, es ist nicht generell möglich, die ursprüngliche Nachricht oder einen Prozess der Übertragung von Signalen wiederempfangenen Signals Unsicherheit

`` The fundamental problem of communication points in the sound completely or select a different item on the news. If the channel is noisy, it is generally not possible, the original message or a process of re-transmission of signals received signal uncertainty.''

Relative Entropy

- The **Relative Entropy** between two distributions p(X) and q(X) is defined as the expectation of the logarithm of the ratio of the distributions

$$D[p(X)||q(X)] = E_p[\log(p(X)/q(X))]$$

Relative Entropy

$$D\left[p(X)\|q(X)\right] = \sum_{i=1}^{N} p(x_i) \;\; \log_b \left[\frac{p(x_i)}{q(x_i)}\right]$$

$$\boxed{D\left[p(XY)\|p(X)p(Y)\right] = \sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j)\ \log_b \left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right] = I(X;Y)}$$