Communication Theory
English Grades

Chapter 4

Fundamental limits in digital communications (Information Theory)

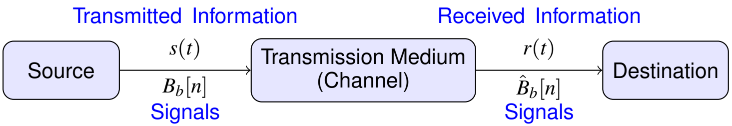
Marcelino Lázaro

Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

Table of Contents

Introduction

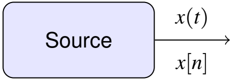
- Purpose of a communications system:
  - *Transmission of information*



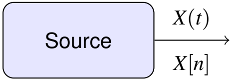
- Information theory
  - Quantitative measures of information
  - Analysis of a communications system
    - ★ Amount of generated information
    - ★ Amount of information that is effectively transmitted (received)
    - ★ Fundamental limits in the transmission of information
- Organization of the chapter:
  - (Probabilistic) models for information sources
  - (Probabilistic) models of the system (channels)
  - Quantitative measures of information
  - Fundamental limits in a digital communication system

Modeling of information sources

- Source output: signal (information flow)
  - Continuous time: $x(t) \equiv s(t)$
  - Discrete-time: $x[n] \equiv B_b[n]$ (or $A[n]$)



- Source output model (information)
  - Random process, $X(t)$, or $X[n]$

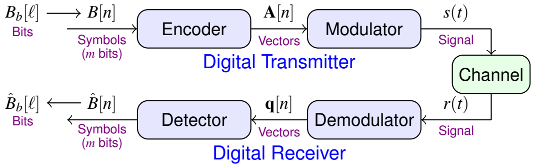


- Model for continuous time (analog) sources
  - Random process $X(t)$ (continuous-time)
    - ★ Characterization: $m_X(t)$, $R_X(t+\tau, t)$, $S_X(j\omega)$
      - Typically they are band limited processes (B Hz)
      - $S_X(j\omega)$ reflects the mean spectral behavior of the source
- Discrete-time source model
  - Random process $X[n]$ (discrete-time)
  - Font alphabet types (possible values of $X[n]$)
    - ★ Continuous (e.g., sampled signals)
    - ★ Discrete (digital sources)
      - Simplest model: discrete memoryless source

Discrete Memoryless Source (DMS)

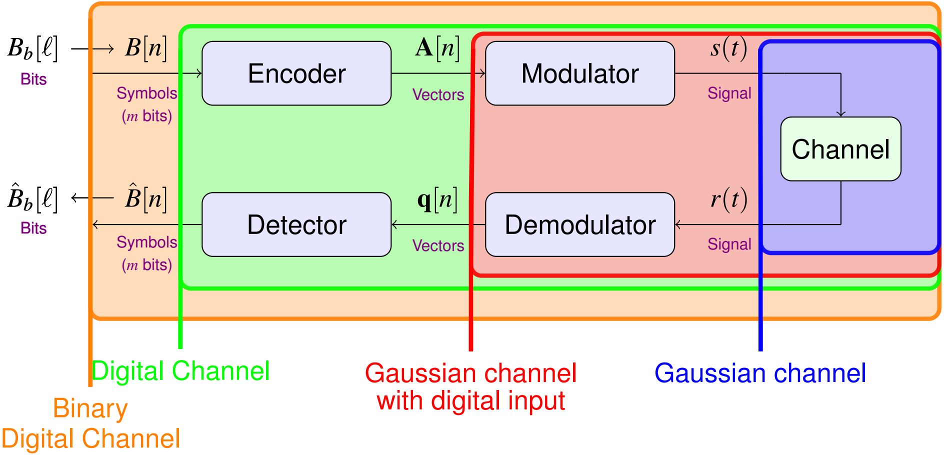
- Source: Discrete-time random process $X[n]$
  - Discrete
    - ★ $X[n]$ alphabet: discrete, of M_X values $\mathcal{A}_X = \{x_0, x_1, \cdots, x_{M_X-1}\}$
  - Memoryless
    - ★ $p_{X[n]}(x_i)$ does not depend on n : $p_{X[n]}(x_i) \equiv p_X(x_i)$
      - $\{X[n]\}$: independent and identically distributed (i.i.d.) R.V.'s
- Description of the process (statistical characterization)
  - Random variable X
    - ★ Being $X[n]$ i.i.d., the statistics are the same for all n
    - ★ Alphabet $\mathcal{A}_X = \{x_0, x_1, \cdots, x_{M_X-1}\}$
    - ★ Probabilities $\{p_X(x_i)\}_{i=0}^{M_X-1} = \{p_X(x_0), p_X(x_1), \cdots, p_X(x_{M_X-1})\}$
- Example: Binary Symmetric Source (BSC)
  - Alphabet $\mathcal{A}_X = \{x_0, x_1\}$, typically $x_0 \equiv 0$, $x_1 \equiv 1$
  - Probabilities $p_X(x_0) = p_X(x_1) = \frac{1}{2}$

Digital Systems - Probabilistic Channel Models

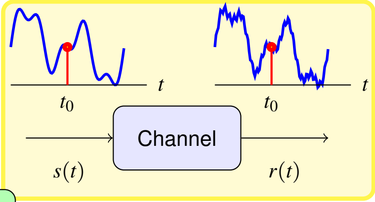


- Probabilistic channel models
  - Probabilistic relationship between the received and transmitted information at different points of this communication model
    - Different levels of abstraction in the definition of *channel*
    - Probabilistic model: Input X, output Y, distribution $f_{Y|X}(y|x)$
- Channel models
  - Gaussian channel: $Y \equiv r(t) \mid X \equiv s(t)$
    - Represents the physical channel
  - Gaussian channel with digital input: $Y \equiv \mathbf{q}[n] \mid X \equiv \mathbf{A}[n]$
    - Represents the equivalent discrete channel
  - Digital channel: $Y \equiv \hat{B}[n] \mid X \equiv B[n]$
    - Represents the transmission of symbols
  - Binary digital channel: $Y \equiv \hat{B}_b[\ell] \mid X \equiv B_b[\ell]$
    - Represents the transmission of bits

Probabilistic Channel Models - Representation



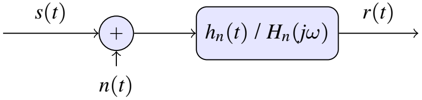
Gaussian channel



- Input / output relationship
  - Input: $X \equiv s(t)$, for a given instant t_0
  - Output: $Y \equiv r(t)$, for the same instant t_0
- Gaussian channel model

$$r(t) = s(t) + n(t)$$

  - $n(t)$: e.g. stationary, white and Gaussian $m_n = 0,\ S_n(j\omega) = \frac{N_0}{2},\ R_n(\tau) = \frac{N_0}{2}\ \delta(\tau)$
- Noise power limitation - Filtering at the input of the receiver



  - Bandwidth of the signal $s(t)$: B Hz ($W = 2\pi B$ rad/s)
    - ★ Ideal $h_n(t)$ filter, bandwidth B Hz: noise power $N_0\ B$ Watts
- Distribution of $Y|X$ when $Y \equiv r(t_0)$ and $X \equiv s(t_0)$)
  - Distribution of Y when $X = s(t_0) \equiv x$
  - Gaussian: mean $x = s(t_0)$ and variance $\sigma^2 = N_0\ B$ $f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-x)^2}{2\sigma^2}}$

Gaussian channel with digital input

- Input / output relationship
  - Input: $\mathbf{X} \equiv \mathbf{A}[n]$, for a given instant n
    - Vector of N random variables (discrete alphabet, $M = 2^m$ values)
    - if $\mathbf{A}[n] = \mathbf{a}_i \rightarrow \mathbf{X} = \mathbf{x}_i \equiv \mathbf{a}_i$
  - Output: $\mathbf{Y} \equiv \mathbf{q}[n]$, for the same instant n
    - Vector of N random variables (continuous alphabet)
- It is equal to the equivalent discrete channel (Chapter 3)
  - It is given the name *Gaussian Channel* in the field of Information Theory (IT)
- Gaussian channel model with digital input

$$\mathbf{q}[n] = \mathbf{A}[n] + \mathbf{z}[n] \qquad \equiv \qquad \mathbf{q} = \mathbf{A} + \mathbf{z}$$

  - Independent of the discrete instant n
  - Distribution of the N elements of the noise vector $\mathbf{z} = [z_0, z_1, \cdots, z_{N-1}]^T$
    - Are N R.V.'s: Independent, Gaussian, zero mean, variance $N_0/2$
- Conditional distribution of the output given the input
  - Gaussian distribution (N-dimensional)
    - Media: the transmitted symbol ($\mathbf{x}_i \equiv \mathbf{a}_i$)
    - Variance: $N_0/2$ in each direction of the N-dimensional space

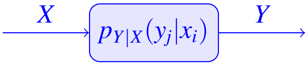
$$f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i) = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{||\mathbf{q}-\mathbf{a}_i||^2}{N_0}} \rightarrow f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_i) = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{||\mathbf{y}-\mathbf{x}_i||^2}{N_0}}$$

Digital Channel

- Input / output relationship
  - Input: symbol for a given instant n
    - $X \equiv B[n]$ (or alternatively $X \equiv \mathbf{A}[n]$)
  - Output: symbol for the same instant n
    - $Y \equiv \hat{B}[n]$ (or alternatively $Y \equiv \hat{\mathbf{A}}[n]$)
  - Alphabet of X and Y : symbols (blocks of m bits / vectors)
    - $M = 2^m$ possible symbols (Alphabet of $B[n]$, or alternatively of $A[n]$)

$$X, Y \in \{x_i = y_i \equiv b_i\}_{i=0}^{M-1} \quad \text{or} \quad X, Y \in \{x_i = y_i \equiv \mathbf{a}_i\}_{i=0}^{M-1}$$

- Probabilistic model: Discrete Memoryless Channel (DMC)



Characterization of the DMC

1. Input alphabet: M_X values $\mathcal{A}_X = \{x_0, x_1, \cdots, x_{M_X-1}\}$
2. Output alphabet: M_Y values $\mathcal{A}_Y = \{y_0, y_1, \cdots, y_{M_Y-1}\}$
3. Set of $M_X \times M_Y$ conditional (transition) probabilities $p_{Y|X}(y_j|x_i)$,

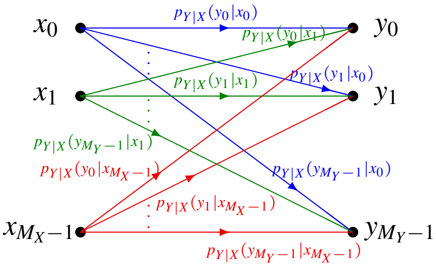
$\forall i \in \{0, 1, \cdots, M_X - 1\}, \ \forall j \in \{0, 1, \cdots, M_Y - 1\}$

DMC: Representation of transition probabilities

- Channel matrix

$$\mathbf{P} = \begin{bmatrix} p_{Y|X}(y_0|x_0) & p_{Y|X}(y_1|x_0) & \cdots & p_{Y|X}(y_{M_Y-1}|x_0) \\ p_{Y|X}(y_0|x_1) & p_{Y|X}(y_1|x_1) & \cdots & p_{Y|X}(y_{M_Y-1}|x_1) \\ \vdots & \vdots & \ddots & \vdots \\ p_{Y|X}(y_0|x_{M_X-1}) & p_{Y|X}(y_1|x_{MX-1}) & \cdots & p_{Y|X}(y_{M_Y-1}|x_{M_X-1}) \end{bmatrix}$$

  - Elements of a row add up to 1
- Arrow (or trellis) diagram



  - Arrows going out of the same node add up to 1

Application of the DMC to the digital channel

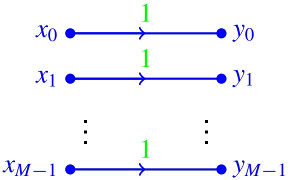
- X and Y alphabets
  - Alphabet of $B[n]$ (m bits): $M = 2^m$ symbols:

$$\mathcal{A}_X = \mathcal{A}_Y = \{b_0, b_1, \cdots, b_{M-1}\}$$

$$x_i \equiv b_i,\ y_j \equiv b_j,\ M_X = M_Y = M,\ i,j \in \{0, 1, \cdots M-1\}$$

- Transition probabilities $p_{Y|X}(y_j|x_i) \equiv p_{\hat{B}|B}(b_j|b_i) = p_{\hat{\mathbf{A}}|\mathbf{A}}(\mathbf{a}_j|\mathbf{a}_i)$
  - Accuracies: $p_{Y|X}(y_i|x_i) = p_{\hat{B}|B}(b_i|b_i)$
  - Error probabilities: $p_{Y|X}(y_j|x_i) = p_{\hat{B}|B}(b_j|b_i)$ for $j \neq i$
  - Ideal values: transition matrix / arrow diagram

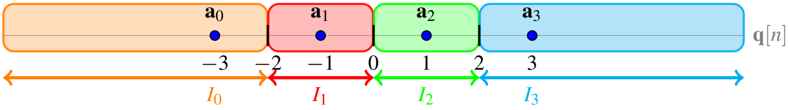
$$\mathbf{P} = \mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & 1 & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$



Calculation of transition probabilities - Example

- $M = 4$, equiprobable symbols $p_A(\mathbf{a}_i) = \frac{1}{4}$
  - Constellation: $\mathbf{a}_0 = -3,\ \mathbf{a}_1 = -1,\ \mathbf{a}_2 = +1,\ \mathbf{a}_3 = +3$
  - Decision regions: thresholds $q_{u1} = -2,\ q_{u2} = 0,\ q_{u3} = +2$

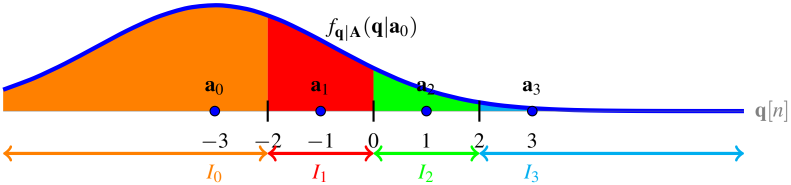
$$I_0 = (-\infty, -2],\ I_1 = (-2, 0],\ I_2 = (0, +2],\ I_3 = (+2, +\infty)$$



- Transition probabilities (channel matrix)

$$\mathbf{P} = \begin{bmatrix} 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) & Q\left(\frac{5}{\sqrt{N_0/2}}\right) \\ Q\left(\frac{1}{\sqrt{N_0/2}}\right) & 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & Q\left(\frac{3}{\sqrt{N_0/2}}\right) \\ Q\left(\frac{3}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) \\ Q\left(\frac{5}{\sqrt{N_0/2}}\right) & Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right) \end{bmatrix}$$

Elements of the first row: $x_0 \equiv \mathbf{a}_0 \rightarrow p_{Y|X}(y_j|x_0), \forall j$



- Distribution $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_0)$: Gaussian with mean $\mathbf{a}_0 = -3$ and variance $N_0/2$

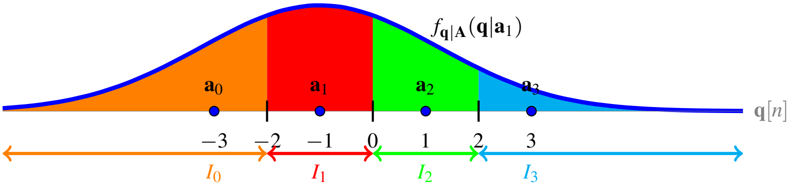
$$p_{Y|X}(y_0|x_0) = P_{a|\mathbf{a}_0} = 1 - P_{e|\mathbf{a}_0} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_1|x_0) = P_{e|\mathbf{a}_0\rightarrow\mathbf{a}_1} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_0) = P_{e|\mathbf{a}_0\rightarrow\mathbf{a}_2} = Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_0) = P_{e|\mathbf{a}_0\rightarrow\mathbf{a}_3} = Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

Elements of the second row: $x_1 \equiv \mathbf{a}_1 \rightarrow p_{Y|X}(y_j|x_1), \forall j$



- Distribution $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_1)$: Gaussian with mean $\mathbf{a}_1 = -1$ and variance $N_0/2$

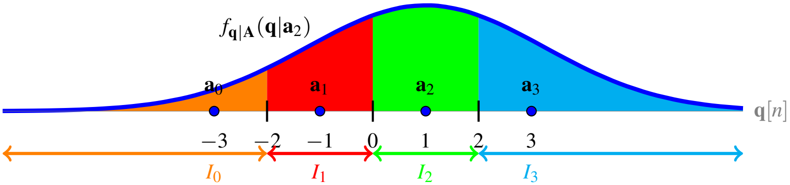
$$p_{Y|X}(y_0|x_1) = P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_0} = Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_1|x_1) = P_{a|\mathbf{a}_1} = 1 - P_{e|\mathbf{a}_1} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_1) = P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_2} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_1) = P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_3} = Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

Third row elements: $x_2 \equiv \mathbf{a}_2 \rightarrow p_{Y|X}(y_j|x_2), \forall j$



- Distribution $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_2)$: Gaussian with mean $\mathbf{a}_2 = +1$ and variance $N_0/2$

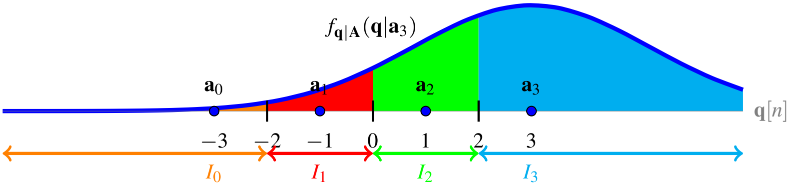
$$p_{Y|X}(y_0|x_2) = P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_0} = Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_1|x_2) = P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_1} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_2) = P_{a|\mathbf{a}_2} = 1 - P_{e|\mathbf{a}_2} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_2) = P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_3} = Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

Fourth row elements: $x_3 \equiv \mathbf{a}_3 \rightarrow p_{Y|X}(y_j|x_3), \forall j$



- Distribution $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_3)$: Gaussian with mean $\mathbf{a}_3 = +3$ and variance $N_0/2$

$$p_{Y|X}(y_0|x_3) = P_{e|\mathbf{a}_3\rightarrow\mathbf{a}_0} = Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

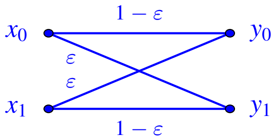
$$p_{Y|X}(y_1|x_3) = P_{e|\mathbf{a}_3\rightarrow\mathbf{a}_1} = Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_3) = P_{e|\mathbf{a}_3\rightarrow\mathbf{a}_2} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_3) = P_{a|\mathbf{a}_3} = 1 - P_{e|\mathbf{a}_3} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

Binary digital channel

- Input/output relationship
  - Input: $X \equiv B_b[\ell]$, bit at instant ℓ
  - Output: $Y \equiv \hat{B}_b[\ell]$, bit at the same instant ℓ
    - Alphabet of X and Y: Bits $x_0 = y_0 \equiv 0$, $x_1 = y_1 \equiv 1$
- Probabilistic model
  - Particular case of the DMC for $M_X = M_Y = 2$
  - 4 conditional probabilities $p_{Y|X}(y_j|x_i)$, para $i,j \in \{0,1\}$
    - 2 success probabilities for bits ($j = i$)
    - 2 error probabilities for bits ($j \neq i$)
- Example: Binary Symmetric Channel
  - Same probability of error for both bits: $p_{Y|X}(y_1|x_0) = p_{Y|X}(y_0|x_1) = \varepsilon$
  - Bit Error Rate (BER): $BER = \varepsilon$
    - $BER = p_X(x_0)\, p_{Y|X}(y_1|x_0) + p_X(x_1)\, p_{Y|X}(y_0|x_1) = \varepsilon$



$$\mathbf{P} = \begin{bmatrix} p_{Y|X}(y_0|x_0) & p_{Y|X}(y_1|x_0) \\ p_{Y|X}(y_0|x_1) & p_{Y|X}(y_1|x_1) \end{bmatrix} = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

Quantitative Measures of Information

Self-information (surprisal) of an event of a discrete random variable

- $I_X(x_i)$: measures the information content of an event of the random variable $(X = x_i)$
- Requirements for such an information measure
  1. Must depend on the probability of the event
     - ★ $I_X(x_i) = f(p_X(x_i))$
  2. Must be a decreasing function of probability
     - ★ $p_X(x_i) > p_X(x_j) \rightarrow I_X(x_i) < I_X(x_j)$
  3. Must be a continuous function of probability
     - ★ $p_X(x_i) \approx p_X(x_j) \rightarrow I_X(x_i) \approx I_X(x_j)$
  4. For a joint event of two independent events $(X = x_i, Y = y_j)$ $(p_{X,Y}(x_i, y_j) = p_X(x_i) \times p_Y(y_j))$
     - ★ $I_{X,Y}(x_i, y_j) = I_X(x_i) + I_Y(y_j)$
- Function that satisfies these properties - Self-information

$$I_X(x_i) = -\log_b(p_X(x_i))$$

  - ▸ The base of the logarithm defines the units of the measurement
    - ★ Base 2 : bits
    - ★ Base e (natural logarithm $\ln$): nats
      NOTE: Relation $\log_b(x) = \ln(x)/\ln(b)$

Entropy (of a discrete R.V.)

- A measure of uncertainty about the outcome of a random variable (information)
  - Alphabet: $\mathcal{A}_X = \{x_0, x_1, \cdots, x_{M_X-1}\}$ (M_X symbols)
  - Probabilities: $\{p_X(x_0), p_X(x_1), \cdots, p_X(x_{M_X-1})\}$
- Average of the self-information of each event

$$H(X) = -\sum_{i=0}^{M_X-1} p_X(x_i)\ \log\left(p_X(x_i)\right) = \sum_{i=0}^{M_X-1} p_X(x_i)\ \log\left(\frac{1}{p_X(x_i)}\right)$$

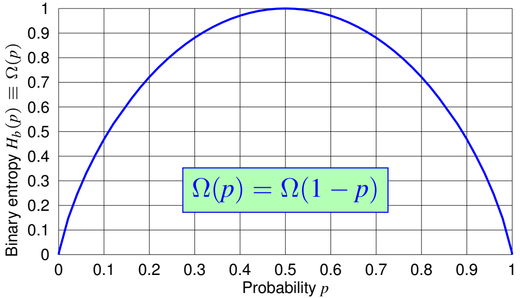
NOTE: By convention: $0 \times \log 0 = 0$

  - Units: bits/symbol (base 2) or nats/symbol (base e)
- Limit values of the entropy of discrete random variables
  1. $H(X) \geq 0$
     - ★ $0 \leq p_X(x_i) \leq 1$ and, consequently, $-\log\left(p_X(x_i)\right) \geq 0$
     - ★ $H(X) = 0$ when $p_X(x_i) = 1,\ p_X(x_j) = 0\ \forall j \neq i$
       - There is no uncertainty about X
  2. $H(X) \leq \log(M_X)$
     - ★ $H(X) = \log(M_X)$ if the symbols are equiprobable $p_X(x_i) = 1/M_X$
       - Situation of maximum uncertainty about X

Example - Binary Entropy: $H_b(p) \equiv \Omega(p)$

- Binary random variable
  - Alphabet: $\{x_0, x_1\}$
  - Probabilities: $\{p_X(x_0) = p, \quad p_X(x_1) = 1-p\}$

$$H(X) \equiv H_b(p) \equiv \Omega(p) = -p\ \log_2(p) - (1-p)\ \log_2(1-p)$$
$$= p\ \log_2\left(\frac{1}{p}\right) + (1-p)\ \log_2\left(\frac{1}{1-p}\right) \text{ bits/symbol}$$



- Maximum value: $\max \Omega(p) = 1$ bit/symbol
  - It is reached at $p = 0.5$ (reference value)

Joint entropy (of two discrete R.V.'s)

- Measure of the joint information of two (or more) random variables

$$H(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \log\left(\frac{1}{p_{X,Y}(x_i,y_j)}\right)$$

- Independent random variables
  - Joint probability: $p_{X,Y}(x_i,y_j) = p_X(x_i)\, p_Y(y_j)$

$$\begin{aligned} H(X,Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_X(x_i)\, p_Y(y_j) \log \frac{1}{p_X(x_i)\, p_Y(y_j)} \\ &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_X(x_i)\, p_Y(y_j) \log \frac{1}{p_X(x_i)} + \sum_{j=0}^{M_Y-1} \sum_{i=0}^{M_X-1} p_X(x_i)\, p_Y(y_j) \log \frac{1}{p_Y(y_j)} \\ &= \sum_{i=0}^{M_X-1} p_X(x_i) \log \frac{1}{p_X(x_i)} \sum_{j=0}^{M_Y-1} p_Y(y_j) + \sum_{j=0}^{M_Y-1} p_Y(y_j) \log \frac{1}{p_Y(y_j)} \sum_{i=0}^{M_X-1} p_X(x_i) \\ &= \sum_{i=0}^{M_X-1} p_X(x_i) \log \frac{1}{p_X(x_i)} + \sum_{j=0}^{M_Y-1} p_Y(y_j) \log \frac{1}{p_Y(y_j)} = H(X) + H(Y) \end{aligned}$$

Conditional entropy (of two discrete R.V.'s)

- Uncertainty in a R.V. when the value of another is known
  - Entropy of X given that $Y = y_j$: $p_X(x_i) \to p_{X|Y}(x_i|y_j)$

$$H(X|Y=y_j) = \sum_{i=0}^{M_X-1} p_{X|Y}(x_i|y_j)\ \log \frac{1}{p_{X|Y}(x_i|y_j)}$$

- Conditional entropy: Average of $H(X|Y=y_j)$
  - Averages over all alphabet values of Y

$$H(X|Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j)\ H(X|Y=y_j)$$

$$= \sum_{j=0}^{M_Y-1} p_Y(y_j) \sum_{i=0}^{M_X-1} p_{X|Y}(x_i|y_j)\ \log \frac{1}{p_{X|Y}(x_i|y_j)}$$

$$H(X|Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_{X|Y}(x_i|y_j)}$$

- For independent random variables
  - Independence: $p_{X|Y}(x_i|y_j) = p_X(x_i) \to H(X|Y=y_j) = H(X), \forall y_j$ $H(X|Y) = H(X)$

Relation between joint and conditional entropy

$$\begin{aligned}
H(X,Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_{X,Y}(x_i,y_j)} \\
&= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_X(x_i)\ p_{Y|X}(y_j|x_i)} \\
&= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_{Y|X}(y_j|x_i)} \\
&= \sum_{i=0}^{M_X-1} p_X(x_i)\ \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_{Y|X}(y_j|x_i)} \\
&= H(X) + H(Y|X)
\end{aligned}$$

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

Mutual information (between two discrete R.V.'s)

- Measures the information provided by a random variable X about the knowledge of another random variable Y

$$I(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \, \log \frac{p_{X,Y}(x_i,y_j)}{p_X(x_i)\, p_Y(y_j)}$$

- Properties

1. $I(X,Y) = I(Y,X) \geq 0$
   - ⋆ The equality holds in the case that X and Y are independent
2. $I(X,Y) \leq \min(H(X),H(Y))$
3. Conditional mutual information can be defined $$I(X,Y|Z) = \sum_{i=0}^{M_Z-1} p_Z(z_i)\, I(X,Y|Z=z_i)$$
   $$I(X,Y|Z) = H(X|Z) - H(X|Y,Z)$$
4. The chain rule for mutual information is $$I((X,Y),Z) = I(X,Z) + I(Y,Z|X)$$
   $$I((X_1,X_2,\cdots,X_N),Y) = I(X_1,Y) + I(X_2,Y|X_1) + \cdots + I(X_N,Y|X_1,\cdots,X_{N-1})$$
5. From the definition of mutual information we obtain the definition of entropy
   $$I(X,X) = H(X)$$

Relationships of mutual information and entropy

$$
\begin{aligned}
I(X,Y) &= \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{p_{X,Y}(x_i,y_j)}{p_X(x_i)\ p_Y(y_j)} \\
&= \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{p_{X|Y}(x_i|y_j)}{p_X(x_i)} \\
&= \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log p_{X|Y}(x_i|y_j) \\
&= \sum_{i=0}^{M_X-1} p_X(x_i)\ \log \frac{1}{p_X(x_i)} - \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j)\ \log \frac{1}{p_{X|Y}(x_i|y_j)} \\
&= H(X) - H(X|Y)
\end{aligned}
$$

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

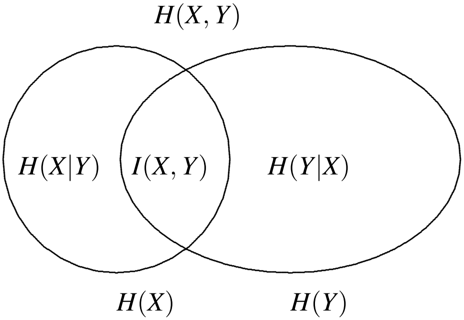
$$H(X,Y) = H(Y) + H(X|Y) \rightarrow H(X|Y) = H(X,Y) - H(Y)$$

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

Relationships of mutual information and entropy (II)

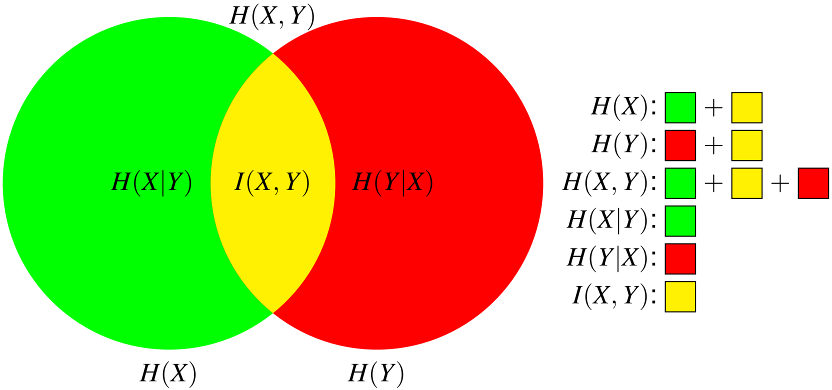
$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$$

- Representation in a Venn diagram
  - Entropies and mutual information represented by areas



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

Venn Diagram - Entropies and Mutual Information



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

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

Differential entropy

- Extension of entropy definitions to continuous random variables

$$h(X) = \int_{-\infty}^{\infty} f_X(x) \ \log \frac{1}{f_X(x)} \ dx$$

Definition of *joint differential entropy*

$$h(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) \log \frac{1}{f_{X,Y}(x,y)} dx\, dy$$

The same is done for the *conditional differential entropy*

$$h(X|Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) \log \frac{1}{f_{X|Y}(x|y)} dx\, dy$$

The alternative but equivalent definition is often used

$$h(X|Y) = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{\infty} f_{X|Y}(x|y) \log \frac{1}{f_{X|Y}(x|y)} dx\, dy$$

Differential Entropy and Mutual Information - Relationships

- Definition of mutual information

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\ \log \frac{f_{X,Y}(x,y)}{f_X(x)\, f_Y(y)} dx\, dy$$

- The same relationships are maintained as for discrete variables

$$h(X,Y) = h(X) + h(Y|X) = h(Y) + h(X|Y)$$

$$I(X,Y) = h(Y) - h(Y|X) = h(X) - h(X|Y) = h(X) + h(Y) - h(X,Y)$$

Differential entropies and mutual information : Properties

- Differential entropies do not maintain the properties of entropies for discrete random variables in terms of interpretation as amount of information
  - Example: $X \sim \mathcal{U}(0,a)$

$$h(X) = \log(a)$$

  - Example: $Y \sim \mathcal{N}(0,\sigma^2)$

$$h(Y) = \frac{1}{2}\log(2\pi\ e\ \sigma^2)$$

  It can be seen that depending on the values of a or σ^2 both entropies can take positive, null or negative values (depending on whether $a \gtreqless 1$, or $\sigma^2 \gtreqless \frac{1}{2\pi e}$)

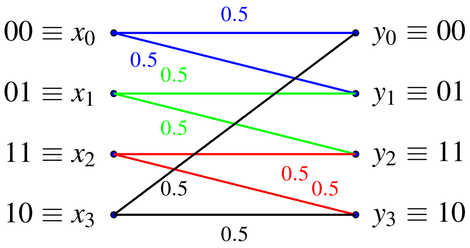
- Mutual information does maintain that intuitive interpretation and the corresponding properties. In particular
  - $I(X,Y) \geq 0$ (non-negative function)
  - $I(X,Y) = 0$ only if X and Y are independent
  - $I(X,Y) = I(Y,X)$

FUNDAMENTAL LIMITS

AT

COMMUNICATIONS SYSTEMS

Reliable transmission on unreliable channels - Example



- 4 symbols $\equiv$ 2 bits of information for channel use
- The channel is unreliable - Errors happen
  - Symbol error probability is $P_e = 1/2$
  - With binary Gray assignment $BER = 1/4$
- Cause of errors
  - Given $Y = y_j$ it is not possible to uniquely identify $X = x_i$
  - Example: $Y = y_0$ is observed
    - The transmitted symbol can be x_0 (no error)
    - The transmitted symbol can be x_3 (with error)

Reliable transmission over unreliable channels - Example (II)

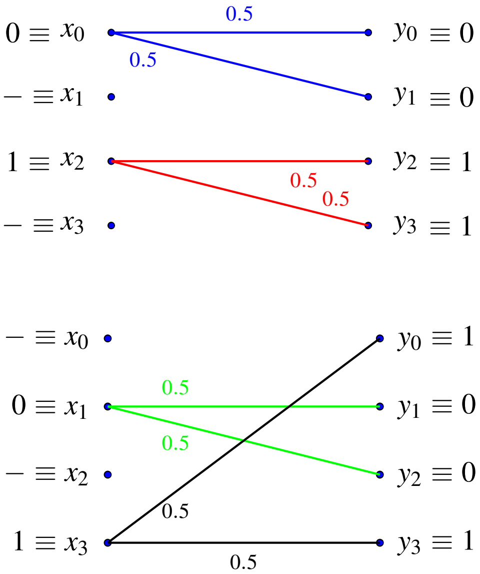
- Option to transmit information reliably
  - To transmit only a subset of the symbols
    - Symbols that generate "*non-overlapping*" outputs
  - Example: transmit only x_0 and x_2
    - x_0: output y_0 or y_1
    - x_2: output y_2 or y_3

  Given an output there is no uncertainty in the transmitted symbol !!!
- It is possible to transmit information on this channel without errors
  - Cost of reliable transmission
    - Less information is transmitted per channel use

      - In this case: 2 symbols $\equiv$ 1 bit for channel use
- Regular channels do not allow this directly
  - Workaround: force similar behavior - Channel coding
    - Zero probability of error is not sought (no overlaps)
    - It seeks to reduce the probability of error arbitrarily

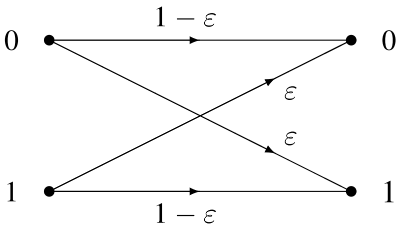
      - Overlaps with arbitrarily low probability

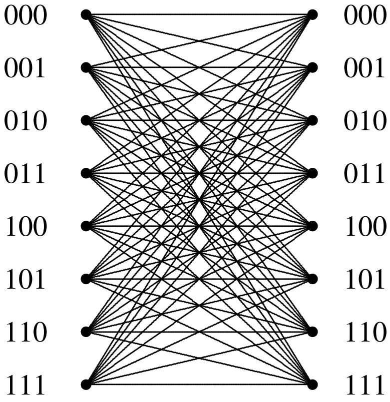
Reliable transmission over unreliable channels - Example (III)



Channel coding

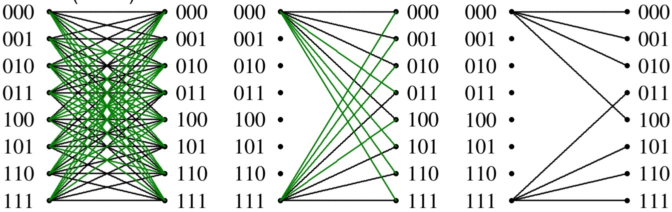
- The channel is used n times together
  - Definition of extended symbols: group of n symbols
- Search for a subset of symbols (2^k) that produce "*low overlapping*" in the output
  - k bits of information are transported for every n uses of the channel
- Example: binary symmetric channel ($BER = \varepsilon$) with $n = 3$





Channel coding (II)

- Most likely situations (for reasonably low ε)
  - 0 errors or 1 error over 3 bits - 4 branches/symbol (in black)
- Less likely situations
  - 2 errors or 3 errors over 3 bits - 4 branches/symbol (in green)
- Subset of 2^k ($k < n$) elements with "*low overlapping*"
  - Example: 000 and 111 ($k = 1$)



- Neglecting "*low probability*" links, there is no overlap
  - Probability of error: $P_e = 3 \times \left[\varepsilon^2\ (1-\varepsilon)\right] + \varepsilon^3$
    - ★ Examples: $\varepsilon = 0.1 \rightarrow P_e = 0.028 \mid \varepsilon = 10^{-3} \rightarrow P_e = 2.998 \times 10^{-6}$
  - Information transmitted: 1 bit (k) of information for every 3 (n) uses of the channel
    - ★ Coding rate: $R = k/n = 1/3$
- Intuition: increasing n and k (with k/n constant) can further reduce
  - There is a limit: Channel capacity

Channel coding (III)

- Example $n = 3$: 8 possible bit triplets (extended symbols)

$$\begin{bmatrix}0\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\end{bmatrix}$$

- Subset of $2^k = 2$ extended symbols with "low overlap"
  - Do not overlap if: there are 0 or 1 bit errors in the transmitted triplet
  - Information transmitted with each triplet: 1 bit ($k = 1$)
  - Coding rate: $R = \frac{k}{n} = \frac{1}{3}$

$$0 \equiv \begin{bmatrix}0\\0\\0\end{bmatrix} \rightarrow \begin{bmatrix}0\\0\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$1 \equiv \begin{bmatrix}1\\1\\1\end{bmatrix} \rightarrow \begin{bmatrix}1\\1\\1\end{bmatrix} \quad \begin{bmatrix}1\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\end{bmatrix}$$

$$P_a = \underbrace{(1-\varepsilon)^3}_{P(0\ err.)} + \underbrace{3 \times \left[\varepsilon^1 (1-\varepsilon)^2\right]}_{P(1\ err.)}, \qquad P_e = \underbrace{3 \times \left[\varepsilon^2 (1-\varepsilon)^1\right]}_{P(2\ err.)} + \underbrace{\varepsilon^3}_{P(3\ err.)}$$

Channel coding (IV)

- Example $n = 6$: 64 possible tuples of bits (extended symbols) - Subset of $2^k = 4$ ext. symbols with “low overlapping”

$$\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\0\\1\end{bmatrix}$$

$$\begin{bmatrix}0\\1\\0\\1\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\1\\0\end{bmatrix}$$

$$\begin{bmatrix}1\\0\\1\\0\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}$$

  - There is no overlap if:
    - ★ There are 0 or 1 erroneous bits in the 6-tuple
    - ★ There are 2 errors in the initial 4 bits of the 6-tuple
    - ★ There are 2 errors: one in the last bit, and another one in the 1st, 2nd or 3rd bit
  - Transmitted information with the 6-tuple: 2 bits ($k = 2$): Coding rate: $R = \frac{k}{n} = \frac{2}{6} = \frac{1}{3}$
  - Probability of success: $P_a = \underbrace{(1-\varepsilon)^6}_{P(0\ err.)} + 6 \times \underbrace{\left[\varepsilon^1(1-\varepsilon)^5\right]}_{P(1\ err.)} + 9 \times \underbrace{\left[\varepsilon^2(1-\varepsilon)^4\right]}_{P(2\ err.)}$

Channel coding (V)

$$00 \equiv \begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} \rightarrow \begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} \quad \begin{bmatrix}1\\0\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix} \quad \begin{bmatrix}1\\1\\0\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix} \quad \begin{bmatrix}1\\0\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\0\\1\end{bmatrix}$$

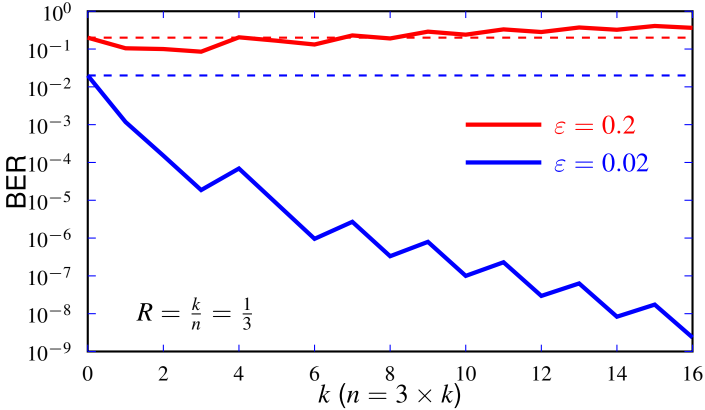
$$01 \equiv \begin{bmatrix}0\\0\\0\\1\\1\\1\end{bmatrix} \rightarrow \begin{bmatrix}0\\0\\0\\1\\1\\1\end{bmatrix} \quad \begin{bmatrix}1\\0\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\1\\0\end{bmatrix} \quad \begin{bmatrix}1\\1\\0\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\1\\1\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\\1\\1\end{bmatrix} \quad \begin{bmatrix}1\\0\\0\\1\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\1\\1\\0\end{bmatrix}$$

$$10 \equiv \begin{bmatrix}1\\1\\1\\1\\0\\1\end{bmatrix} \rightarrow \begin{bmatrix}1\\1\\1\\1\\0\\1\end{bmatrix} \quad \begin{bmatrix}0\\1\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\0\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\\1\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}0\\1\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\0\\1\\0\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\0\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\0\\1\end{bmatrix} \quad \begin{bmatrix}0\\1\\1\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\0\\0\end{bmatrix}$$

$$11 \equiv \begin{bmatrix}1\\1\\1\\0\\1\\0\end{bmatrix} \rightarrow \begin{bmatrix}1\\1\\1\\0\\1\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\\1\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\0\\0\end{bmatrix} \begin{bmatrix}1\\1\\1\\0\\1\\1\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\1\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\\0\\1\\0\end{bmatrix} \begin{bmatrix}1\\0\\1\\1\\1\\0\end{bmatrix} \begin{bmatrix}1\\1\\0\\1\\1\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\\1\\0\\1\\1\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\0\\0\\1\\1\end{bmatrix}$$

$$P_e = (15-9) \times \underbrace{\left[\varepsilon^2(1-\varepsilon)^4\right]}_{P(2\ err.)} + 20 \times \underbrace{\left[\varepsilon^3(1-\varepsilon)^3\right]}_{P(3\ err.)} + 15 \times \underbrace{\left[\varepsilon^4(1-\varepsilon)^2\right]}_{P(4\ err.)} + 6 \times \underbrace{\left[\varepsilon^5(1-\varepsilon)\right]}_{P(5\ err.)} + \underbrace{\varepsilon^6}_{P(6\ err.)}$$

Channel coding (VI)



- Channel Capacity

$\varepsilon = 0.2 \rightarrow C = 0.278 \quad (R > C)$

$\varepsilon = 0.02 \rightarrow C = 0.858 \quad (R < C)$

Channel coding for error detection and correction

- Channel coding: introduction of structured redundancy
  - k information symbols are carried by transmitting n symbols ($n > k$)
    - ★ Coding rate: $R = \frac{k}{n}$
    - ★ Dictionary of the code $\mathcal{C}(k, n)$

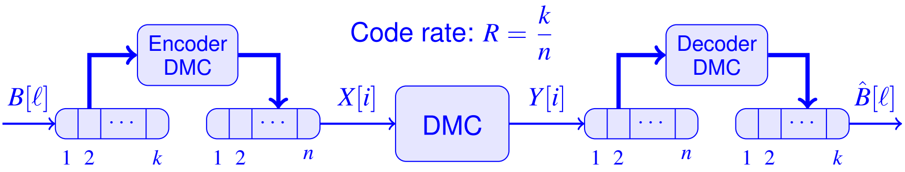
Example of dictionary for two binary codes

Index set (k)	Code words (n)
0	000
1	111

Example code $\mathcal{C}(1,3)$

Index set (k)	Code words (n)
00	000000
01	000111
10	111101
11	111010

Example code $\mathcal{C}(2,6)$



Noisy-Channel Coding Theorem

Channel Capacity: C

Noisy-Channel Coding Theorem (Claude Shannon 1948):

1. If the transmission rate R is less than C, then for any $\delta > 0$ there exists a code with block length n long enough whose probability of error is less than δ
   - Channel coding: allows the probability of error to be reduced to any arbitrarily low level
2. If $R > C$, the error probability of any code with any block length n is limited by a non-null value
   - Channel coding: **DOES NOT** allow the probability of error to be reduced to any arbitrarily low level
3. There are codes that allow reaching the channel capacity $R = C$

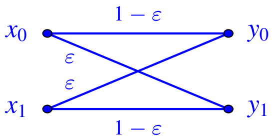
Channel capacity

- Maximum amount of information that can be transmitted reliably through a communications channel in a digital communications system
  - Distortion occurs in the transmission
    - Potential loss of information
  - Reliable Transmission - Definition
    - Ideal: transmission without potential loss of information
    - In practice: transmission capable of reducing the probability of error as much as necessary
  - Channel coding concept
    - Mechanism that allows a reliable transmission
  - Channel capacity
    - Limit on the number of extended symbols with arbitrarily low overlap as the number of channel uses tends to infinity
- Study of channel capacity:
  - Digital channel (DMC)
    - The binary digital channel is considered a particular case
  - Gaussian channel

DMC: Channel capacity through mutual information

- Mutual information between the input and output of a DMC

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



- Analysis for a BSC with $BER = \varepsilon$ in two extreme cases
  - Optimal case (ideal channel): $\varepsilon = 0$
    $H(X|Y) = 0 \quad \rightarrow \quad I(X,Y) = H(X)$
  - Worst case: $\varepsilon = 1/2$
    $H(X|Y) = H(X) \quad \rightarrow \quad I(X,Y) = 0$ — X e Y independent
- The following conclusions can be drawn
  1. The mutual information between input and output of the channel can be see as the amount of information that passes from the input to the output when using channel
     - ★ In an ideal channel ($\varepsilon = 0$) all the information passes: $I(X,Y) = H(X)$
     - ★ If input and output are independent, no information passes: $I(X,Y) = 0$
  2. $H(X|Y)$ can be interpreted as the information that is "*lost*" in the channel
     - ★ In an ideal channel ($\varepsilon = 0$) the loss is null: $H(X|Y) = 0$
     - ★ If input and output are independent, the loss is total: $H(X|Y) = H(X)$
     - ★ $I(X,Y) = H(X) - H(X|Y)$: information at the input, minus the information that is lost

Channel capacity for a digital channel

- Formal definition for a DMC

$$C = \max_{p_X(x_i)} I(X,Y)$$

  - Its units are bits (or bits per channel use)
  - Maximization over $p_X(x_0), p_X(x_1), \cdots, p_X(x_{M_X-1})$
- Limit values

$$0 \leq C \leq \min\{\log M_X, \log M_Y\}$$

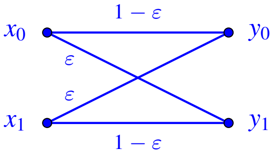
  - $C \geq 0$
    - Since $I(X,Y) \geq 0$
  - $C \leq \log M_X$
    - Since $I(X,Y) \leq H(X)$, and $H(X) \leq \log M_X$
  - $C \leq \log M_Y$
    - Since $I(X,Y) \leq H(Y)$, and $H(Y) \leq \log M_Y$

Capacity: constrained maximization problem

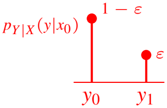
$$C = \max_{p_X(x_i)} I(X,Y)$$

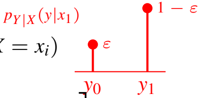
- Approach: maximization of a function with restrictions
  - Function to maximize
    - $I(X,Y)$
  - Variables over which it is maximized (unknowns)
    - $p_X(x_0), p_X(x_1), \cdots, p_X(x_{M_X-1})$
  - Constraints (on the unknowns)
    - $0 \leq p_X(x_i) \leq 1$, for $i \in \{0, 1, \cdots, M_X - 1\}$
    - $\sum_{i=0}^{M_X-1} p_X(x_i) = 1$
- In general, finding an analytical solution can be difficult
  - Analytical solutions can be obtained only for "*simple*" channels
  - Calculation by numerical methods using computers

Binary Symmetric Channel (BSC) : $BER = \varepsilon$



- Capacity: $C = \max_{p_X(x_i)} I(X,Y)$
- Calculation of $I(X,Y)$





$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^{1} p_X(x_i)\; H(Y|X=x_i)$$

$$= H(Y) - \sum_{i=0}^{1} p_X(x_i) \underbrace{\left[-\sum_{j=0}^{1} p_{Y|X}(y_j|x_i)\; \log p_{Y|X}(y_j|x_i) \right]}_{-\varepsilon\;\log(\varepsilon)-(1-\varepsilon)\;\log(1-\varepsilon)=\Omega(\varepsilon)} = H(Y) - \Omega(\varepsilon)$$

- Calculation of channel capacity
  - The maximum of mutual information is sought
    - ★ For this channel $H(Y|X) = \Omega(\varepsilon)$ does not depend on $p_X(x_i)$
      - $I(X,Y)$ is maximum when $H(Y)$ is maximum
    - ★ $H(Y)$ is maximum if $p_Y(y_j) = \frac{1}{M_Y}$: $\max H(Y) = \log M_Y = 1$ bit/symb.

$$\left.\begin{array}{l} p_Y(y_0) = \frac{1}{2} = p_X(x_0)(1-\varepsilon) + p_X(x_1)\;\varepsilon \\ p_Y(y_1) = \frac{1}{2} = p_X(x_0)\;\varepsilon + p_X(x_1)(1-\varepsilon) \end{array}\right\} \rightarrow p_X(x_0) = \frac{1}{2},\; p_X(x_1) = \frac{1}{2}$$

$C = 1 - \Omega(\varepsilon)$ bits/use

$p_X(x_0) = p_X(x_1) = \frac{1}{2}$

Limits for transmission on a digital channel

- A digital channel has a capacity of C bits/use
  - If channel codes are used, the practical codes (those that make it possible to reduce the probability of error arbitrarily) must have a coding rate lower than C

$$R < C$$

- Practical limitation in terms of effective transmission rate when encoding for error protection is used
  - System designed to transmit at R_b bits/s (raw rate)
    - ⋆ Effective rate:

$$R_b^{effective} = R \times R_b \text{ information bits/s}$$

    - ⋆ Effective transmission rate limit

$$R_b^{effective} < C \times R_b \text{ bits of information/s}$$

Channel capacity for Gaussian channel

- Input-output relationship model in a Gaussian channel

$$Y = X + Z$$

Z is a Gaussian random variable, with zero mean and variance P_Z

- Channel capacity under the following conditions:
  - Transmitted power: P_X watts
  - Bandwidth: B Hz
    - Noise power: $P_Z = N_0\, B$ watts
- Calculation through mutual information $C = \max_{f_X(x)\,|\,E[X^2] \leq P_X} I(X,Y)$

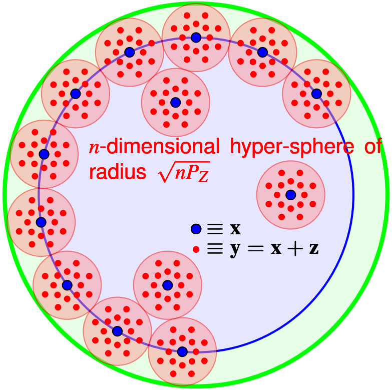
Constraint: $E[X^2] \leq P_X$ given by the power limitation

- Result

$$C = B\, \log_2 \left(1 + \frac{P_X}{N_0 B}\right) \text{ bits/s}$$

It is obtained for a Gaussian $f_X(x)$

Channel capacity for Gaussian channel (II)



n-dimensional hyper-sphere: radius $\sqrt{nP_X}$

n-dim. hyper-sphere: radius $\sqrt{n(P_X + P_Z)}$

Gaussian channel capacity under:

- Transmitted power: P_X watts
- Bandwidth: B Hz
  - Noise power: $P_Z = N_0 B$ watts

Capacity: number of non-overlapping spheres for n uses given (taking into account the noise)

$$M_{no} = \left(1 + \frac{P_X}{P_Z}\right)^{n/2} \quad C = \frac{\log_2 M_{no}}{n} = \frac{1}{2} \log_2 \left(1 + \frac{P_X}{P_Z}\right)$$

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P_X}{N_0 B}\right) \text{ bits/use}$$

Number of transmissions/sec: $2B$

$$C = B \log_2 \left(1 + \frac{P_X}{N_0 B}\right) \text{ bits/s}$$

Capacity of the Gaussian channel - Effect of P_X and B

- Capacity depends on two design parameters $C = B \log_2\left(1 + \frac{P_X}{N_0 B}\right)$ bits/s
  - Power of the transmitted signal, P_X
  - Available bandwidth in Hz, B
- Channel capacity as a function of transmitted power P_X

$$\lim_{P_X \to \infty} C = \infty$$

  - C can be arbitrarily increased by increasing P_X
  - Linear increase of C requires exponential increase of P_X
- Channel capacity as a function of bandwidth (B Hz)

$$\lim_{B \to \infty} C = \frac{P_X}{N_0} \log_2(e) = 1.44 \frac{P_X}{N_0}$$

  - The increment of B does not allow an arbitrary increment of C

Limits for transmission in a Gaussian channel

Binary transmission rate

R_b bits/s $< C$

Signal-to-noise ratio

$SNR = \frac{P_X}{P_Z} = \frac{P_X}{N_0 B}$

$C = B \log_2\left(1 + \frac{P_X}{N_0 B}\right)$ bits/s

$\eta = \frac{R_b}{B}$ bits/s/Hz

Spectral binary rate
(Spectral Efficiency)

$E_b = \frac{P_X}{R_b}$ J/bit

Average energy per bit

$\frac{E_b}{N_0} = \frac{P_X}{R_b N_0} = \frac{P_X}{N_0 B} \times \frac{B}{R_b} = \frac{SNR}{\eta}$

E_b / N_0 ratio

- Practical communication system $R_b < C \rightarrow R_b < B \log_2(1 + SNR)$ bits/s
  - Dividing by B on both sides and rearranging

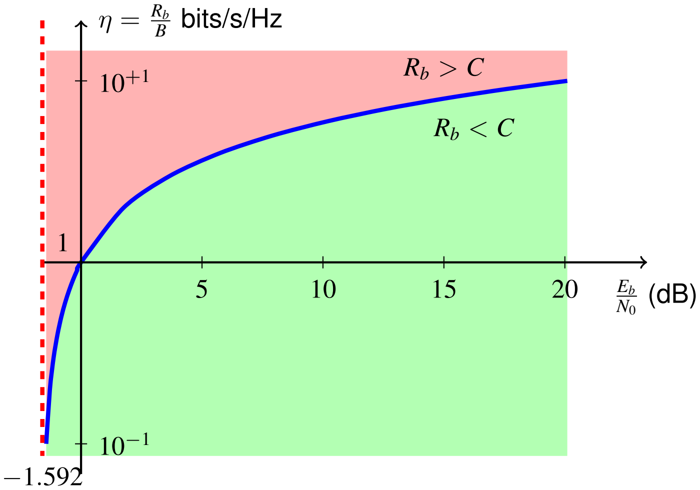
$$\eta < \log_2(1 + SNR), \qquad \eta < \log_2\left(1 + \eta \frac{E_b}{N_0}\right)$$

$$SNR > 2^\eta - 1, \qquad \frac{E_b}{N_0} > \frac{2^\eta - 1}{\eta}$$

If $\eta \rightarrow 0$ then $\frac{E_b}{N_0} = \ln 2 = 0.693 \approx -1.6$ dB

Spectral Binary Rate vs. E_b/N_0

- The curve is represented on the plane η vs $\frac{E_b}{N_0}$ $\frac{E_b}{N_0} = \frac{2^\eta - 1}{\eta}$
  - Divide the plane into two regions: systems with $R_b < C$ (practical) and with $R_b > C$



Normalized signal-to-noise ratio

- Lower bound for SNR

$$SNR > 2^\eta - 1$$

- Definition of normalized SNR

$$SNR_{norm} = \frac{SNR}{2^\eta - 1}$$

- Lower bound on SNR_{norm}

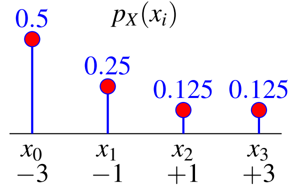
$$SNR_{norm} > 1 \text{ (0 dB)}$$

EXAMPLES

OF

INFORMATION MEASUREMENTS

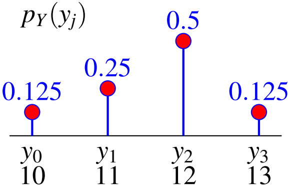
Example - Entropy



x_i	$x_0=-3$	$x_1=-1$	$x_2=+1$	$x_3=+3$
$p_X(x_i)$	$1/2$	$1/4$	$1/8$	$1/8$

$$H(X) = -\sum_{i=0}^{M_X-1} p_X(x_i)\log_2 p_X(x_i) = \sum_{i=0}^{M_X-1} p_X(x_i)\log_2 \frac{1}{p_X(x_i)}$$

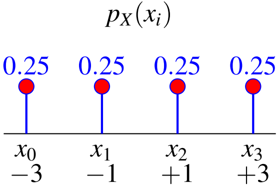
$$H(X) = \frac{1}{2}\underbrace{\log_2 2}_{1} + \frac{1}{4}\underbrace{\log_2 4}_{2} + \frac{1}{8}\underbrace{\log_2 8}_{3} + \frac{1}{8}\underbrace{\log_2 8}_{3} = \frac{7}{4} \text{ bits/symb.}$$



y_j	$y_0=10$	$y_1=11$	$y_2=12$	$y_3=13$
$p_Y(y_j)$	$1/8$	$1/4$	$1/2$	$1/8$

$$H(Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j)\log_2 \frac{1}{p_Y(x_j)} = ?$$

Example - Entropy : Equiprobable symbols



x_i	$x_0=-3$	$x_1=-1$	$x_2=+1$	$x_3=+3$
$p_X(x_i)$	$1/4$	$1/4$	$1/4$	$1/4$

$$H(X) = -\sum_{i=0}^{M_X-1} p_X(x_i)\log_2 p_X(x_i) = \sum_{i=0}^{M_X-1} p_X(x_i)\log_2 \frac{1}{p_X(x_i)}$$

$$H(X) = \frac{1}{4}\underbrace{\log_2 4}_{2} + \frac{1}{4}\underbrace{\log_2 4}_{2} + \frac{1}{4}\underbrace{\log_2 4}_{2} + \frac{1}{4}\underbrace{\log_2 4}_{2} = 2 \text{ bits/symb.}$$

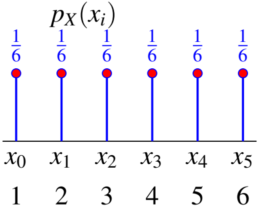
- Maximum value of the entropy

$$H(X) \leq \log_2 M_X$$

  - Achieved for equiprobable symbols
    - Situation of maximum uncertainty

Example - Die : Entropy

$$\mathcal{A}_X = \{1, 2, 3, 4, 5, 6\}$$

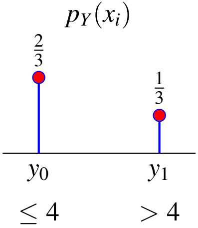


x_i	1	2	3	4	5	6
$p_X(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$H(X) = \log_2 6 = 2.585$ bits/symb.

Maximum uncertainty !!!

$$\mathcal{A}_Y = \{\leq 4, > 4\}$$



y_j	≤ 4	> 4
$p_Y(y_j)$	$\frac{2}{3}$	$\frac{1}{3}$

$$\begin{aligned} H(Y) &= -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \\ &= \frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2 3 \\ &= \Omega\left(\frac{2}{3}\right) = \Omega\left(\frac{1}{3}\right) = 0.918 \text{ bits/symb.} \end{aligned}$$

Lower uncertainty !!!

Example - Die : Joint Entropy

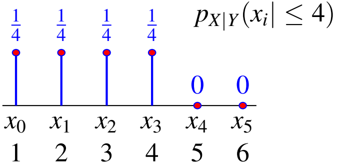
$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/6	1/6	1/6	1/6	0	0	2/3
> 4	0	0	0	0	1/6	1/6	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	

$$H(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X,Y}(x_i, y_j)}$$

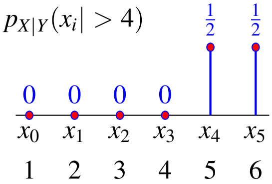
$$\begin{aligned} H(X,Y) &= \frac{1}{6}\log_2 6 + \frac{1}{6}\log_2 6 + \cdots \\ &= 6 \times \frac{1}{6}\log_2 6 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_{0} = \log_2 6 = 2.585 \text{ bits/symb.} \end{aligned}$$

$$H(X,Y) = 2.585 \neq H(X) + H(Y) = 3.503 \text{ bits/symb.}$$

Example - Die : Conditional Entropy $H(X|Y)$



$$H(X|Y = \leq 4) = 4 \times \tfrac{1}{4}\log_2 4$$
$$= \log_2 4 = 2 \text{ bits/symb.}$$



$$H(X|Y = > 4) = 2 \times \tfrac{1}{2}\log_2 2$$
$$= \log_2 2 = 1 \text{ bits/symb.}$$

Half of the uncertainty !!!

$$H(X|Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j) H(X|Y = y_j) = \frac{2}{3} \times 2 + \frac{1}{3} \times 1 = \frac{5}{3} \text{ bits/symb.}$$

Lower than $H(X) = \log_2 6 = 2.585$ bits/symb.

Example - Die : Conditional Entropy $H(X|Y)$

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$1/6$	$1/6$	$1/6$	$1/6$	0	0	$2/3$
> 4	0	0	0	0	$1/6$	$1/6$	$1/3$
$p_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$2/3$

$$p_{X|Y}(x_i|y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)}$$

$p_{X|Y}(x_i|y_j)$	1	2	3	4	5	6	$\sum_i p_{X|Y}(x_i|y_j)$
≤ 4	$1/4$	$1/4$	$1/4$	$1/4$	0	0	1
> 4	0	0	0	0	$1/2$	$1/2$	1

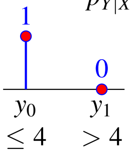
$$H(X|Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X|Y}(x_i|y_j)} = \frac{1}{6} \log_2 4 + \frac{1}{6} \log_2 4 + \cdots$$

$$= 4 \times \frac{1}{6} \log_2 4 + 2 \times \frac{1}{6} \log_2 2 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_{0}$$

$$= \frac{2}{3} \log_2 4 + \frac{1}{2} \log_2 2 = \frac{5}{3} \text{ bits/symb.}$$

Example - Die : Conditional Entropy $H(Y|X)$

$p_{Y|X}(y_j|1)$

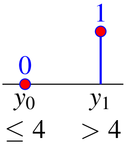


$$H(Y|X=1) = -1\log_2 1 - 0\log_2 0$$
$$= 0 \text{ bits/symb.}$$

No uncertainty !!!

$$p_{Y|X}(y_j|1) = p_{Y|X}(y_j|2) = p_{Y|X}(y_j|3) = p_{Y|X}(y_j|4)$$
$$H(Y|X=1) = H(Y|X=2) = H(Y|X=3) = H(Y|X=4) = 0 \text{ bits/symb.}$$

$p_{Y|X}(y_j|5) = p_{Y|X}(y_j|6)$



$$H(Y|X=5) = H(Y|X=6) = -1\log_2 1 - 0\log_2 0$$
$$= 0 \text{ bits/symb.}$$

No uncertainty !!!

$$H(Y|X) = 0 \text{ bits/symb.}$$

Example - Die : Conditional Entropy $H(Y|X)$

$p_{X,Y}(x_i,y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$1/6$	$1/6$	$1/6$	$1/6$	0	0	$2/3$
> 4	0	0	0	0	$1/6$	$1/6$	$1/3$
$p_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$2/3$

$$p_{Y|X}(y_j|x_i) = \frac{p_{X,Y}(x_i,y_j)}{p_X(x_i)}$$

$p_{Y|X}(y_j|x_i)$	1	2	3	4	5	6
≤ 4	1	1	1	1	0	0
> 4	0	0	0	0	1	1
$\sum_j p_{Y|X}(y_j|x_i)$	1	1	1	1	1	1

$$H(Y|X) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \log_2 \frac{1}{p_{Y|X}(y_j|x_i)}$$

$$= 6 \times \frac{1}{6} \log_2 1 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_{0} = 0 \text{ bits/symb.}$$

Example - Die : Mutual Information $I(X,Y)$

$p_{X,Y}(x_i,y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/6	1/6	1/6	1/6	0	0	2/3
> 4	0	0	0	0	1/6	1/6	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

$$I(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \log_2 \frac{p_{X,Y}(x_i,y_j)}{p_X(x_i)\, p_Y(y_j)}$$

$$\begin{aligned} I(X,Y) &= \frac{1}{6} \log_2 \frac{\frac{1}{6}}{\frac{1}{6} \times \frac{2}{3}} + \ldots + \frac{1}{6} \log_2 \frac{\frac{1}{6}}{\frac{1}{6} \times \frac{1}{3}} \\ &= 4 \times \frac{1}{6} \log_2 \frac{3}{2} + 2 \times \frac{1}{6} \log_2 3 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_{0} \end{aligned}$$

$$I(X,Y) = \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3$$

$$I(X,Y) = \Omega\left(\frac{2}{3}\right) = \Omega\left(\frac{1}{3}\right) = 0.9183 \text{ bits/symb.}$$

Relationships between quantitative information measures

$H(X) = \log_2 6 = 2.585$ bits/symb.

$H(Y) = \Omega(1/3) = 0.918$ bits/symb.

$H(X|Y) = \frac{5}{3}$ bits/symb.

$H(Y|X) = 0$ bits/symb.

$H(X,Y) = \log_2 6 = 2.585$ bits/symb.

$I(X,Y) = \Omega(1/3) = 0.918$ bits/symb.

- Relationships for joint entropy

$$H(X,Y) = H(X) + H(Y|X) = \log_2 6 + 0 = \log_2 6 = 2.585 \text{ bits/symb.}$$

$$H(X,Y) = H(Y) + H(X|Y) = \Omega(1/3) + \frac{5}{3} = 2.585 \text{ bits/symb.}$$

$$H(X,Y) \neq H(X) + H(Y) = 3.503 \text{ bits/symb. (NOT Independent!!!)}$$

Relationships between quantitative information measures (II)

$H(X) = \log_2 6 = 2.585$ bits/symb. $\quad$ $H(Y) = \Omega(1/3) = 0.918$ bits/symb.

$H(X|Y) = \frac{5}{3}$ bits/symb. $\quad$ $H(Y|X) = 0$ bits/symb.

$H(X,Y) = \log_2 6 = 2.585$ bits/symb. $\quad$ $I(X,Y) = \Omega(1/3) = 0.918$ bits/symb.

- Relationships for mutual information

$$I(X,Y) = H(Y) - H(Y|X) = \Omega(1/3) - 0 = 0.918 \text{ bits/symb.}$$

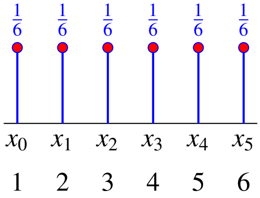
$$I(X,Y) = H(X) - H(X|Y) = \log_2 6 - \frac{5}{3} = 0.918 \text{ bits/symb.}$$

$$I(X,Y) = H(X) + H(Y) - H(X,Y) = \log_2 6 + \Omega(1/3) - \log_2 6 = 0.918 \text{ bits/symb.}$$

$I(X,Y) \neq 0$ bits/symb. (NOT Independent!!!)

Example Two Dice - Entropy

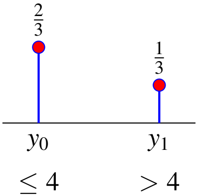
Die 1 : $p_X(x_i)$



x_i	1	2	3	4	5	6
$p_X(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$H(X) = \log_2 6 = 2.585$ bits/symb.

Die 2 : $p_Y(x_i)$



y_j	≤ 4	> 4
$p_Y(y_j)$	$\frac{2}{3}$	$\frac{1}{3}$

$$\begin{aligned} H(Y) &= -\tfrac{2}{3}\log_2\tfrac{2}{3} - \tfrac{1}{3}\log_2\tfrac{1}{3} \\ &= \tfrac{2}{3}\log_2\tfrac{3}{2} + \tfrac{1}{3}\log_2 3 \\ &= \Omega\left(\tfrac{2}{3}\right) = \Omega\left(\tfrac{1}{3}\right) = 0.918 \text{ bits/symb.} \end{aligned}$$

Example Two Dice - Entropy conjunta

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$2/3$
> 4	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/3$
$p_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$2/3$

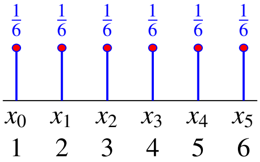
$$H(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X,Y}(x_i, y_j)}$$

$$= 6 \times \frac{1}{9} \log_2 9 + 6 \times \frac{1}{18} \log_2 18 = 3.503 \text{ bits/symb.}$$

X and Y are INDEPENDENT

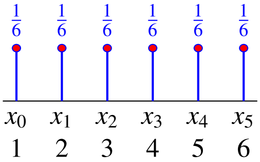
$$H(X,Y) = H(X) + H(Y) = 2.585 + 0.918 = 3.503 \text{ bits/symb.}$$

Example Two Dice - Conditional Entropy $H(X|Y)$



$p_{X|Y}(x_i| \leq 4)$

$$H(X|Y = \leq 4) = H(X)$$
$$= \log_2 6 = 2.585 \text{ bits/symb.}$$



$p_{X|Y}(x_i| > 4)$

$$H(X|Y = > 4) = H(X)$$
$$= \log_2 6 = 2.585 \text{ bits/symb.}$$

X and Y are INDEPENDENT

$$H(X|Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j) H(X|Y = y_j) = H(X) = \log_2 6 = 2.585 \text{ bits/symb.}$$

Example Two Dice - Conditional Entropy $H(X|Y)$

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$2/3$
> 4	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/3$
$p_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$2/3$

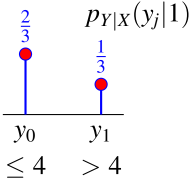
$$p_{X|Y}(x_i|y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)}$$

$p_{X|Y}(x_i|y_j)$	1	2	3	4	5	6	$\sum_i p_{X|Y}(x_i|y_j)$
≤ 4	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1
> 4	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

$$H(X|Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X|Y}(x_i|y_j)}$$

$$= 6 \times \frac{1}{9} \log_2 6 + 6 \times \frac{1}{18} \log_2 6 = \log_2 6 \text{ bits/symb.}$$

Example Two Dice - Conditional Entropy $H(Y|X)$



$$H(Y|X=1) = H(Y) = \Omega(1/3) = 0.918 \text{ bits/symb.}$$

$$p_{Y|X}(y_j|1) = p_{Y|X}(y_j|2) = p_{Y|X}(y_j|3) = p_{Y|X}(y_j|4) = p_{Y|X}(y_j|5) = p_{Y|X}(y_j|6)$$

$$H(Y|X=1) = H(Y|X=2) = H(Y|X=3) = H(Y) = 0.918 \text{ bits/symb.}$$

$$H(Y|X=4) = H(Y|X=5) = H(Y|X=6) = H(Y) = 0.918 \text{ bits/symb.}$$

X and Y are INDEPENDENT

$$H(Y|X) = \sum_{i=0}^{M_X-1} p_X(x_i) H(Y|X=x_i) = H(Y) = 0.918 \text{ bits/symb.}$$

Example Two Dice - Conditional Entropy $H(Y|X)$

$p_{X,Y}(x_i,y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$2/3$
> 4	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/3$
$p_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$2/3$

$$p_{Y|X}(y_j|x_i) = \frac{p_{X,Y}(x_i,y_j)}{p_X(x_i)}$$

$p_{Y|X}(y_j|x_i)$	1	2	3	4	5	6
≤ 4	$2/3$	$2/3$	$2/3$	$2/3$	$2/3$	$2/3$
> 4	$1/3$	$1/3$	$1/3$	$1/3$	$1/3$	$1/3$
$\sum_j p_{Y|X}(y_j|x_i)$	1	1	1	1	1	1

$$H(Y|X) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \log_2 \frac{1}{p_{Y|X}(y_j|x_i)}$$

$$= 6 \times \frac{1}{9} \log_2 \frac{3}{2} + 6 \times \frac{1}{18} \log_2 3$$

$$= \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3 = \Omega(1/3) = 0.918 \text{ bits/symb.}$$

Example Two Dice - Información mutua $I(X,Y)$

$p_{X,Y}(x_i,y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$2/3$
> 4	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/18$	$1/3$
$p_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$2/3$

$$I(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \log_2 \frac{p_{X,Y}(x_i,y_j)}{p_X(x_i)\, p_Y(y_j)}$$
$$= \frac{1}{9} \log_2 \frac{\frac{1}{9}}{\frac{1}{6} \times \frac{2}{3}} + \ldots + \frac{1}{18} \log_2 \frac{\frac{1}{18}}{\frac{1}{6} \times \frac{1}{3}}$$
$$= 6 \times \frac{1}{18} \log_2 1 = 0 \text{ bits/symb.}$$

X and Y are INDEPENDENT

$$I(X,Y) = 0 \text{ bits/symb.}$$

Relationships between quantitative information measures

- Example Two Dice

$H(X) = \log_2 6 = 2.585$ bits/symb. $H(Y) = \Omega\,(1/3) = 0.918$ bits/symb.

$H(X|Y) = \log_2 6 = 2.585$ bits/symb. $H(Y|X) = \Omega\,(1/3) = 0.918$ bits/symb.

$H(X,Y) = 3.503$ bits/symb. $I(X,Y) = 0$ bits/symb.

- Relationships for the joint entropy

$$H(X,Y) = H(X) + H(Y|X) = \log_2 6 + \Omega\,(1/3) = 3.503 \text{ bits/symb.}$$

$$H(X,Y) = H(Y) + H(X|Y) = \Omega(1/3) + \frac{5}{3} = 3.503 \text{ bits/symb.}$$

$$H(X,Y) = H(X) + H(Y) = 3.503 \text{ bits/symb. (INDEPENDENT!!!)}$$

Relationships between quantitative information measures (II)

- Example Two Dice

$$H(X) = \log_2 6 = 2.585 \text{ bits/symb.} \qquad H(Y) = \Omega\,(1/3) = 0.918 \text{ bits/symb.}$$

$$H(X|Y) = \log_2 6 = 2.585 \text{ bits/symb.} \qquad H(Y|X) = \Omega\,(1/3) = 0.918 \text{ bits/symb.}$$

$$H(X,Y) = 3.503 \text{ bits/symb.} \qquad I(X,Y) = 0 \text{ bits/symb.}$$

- Relaitonships for the mutual information

$$I(X,Y) = H(Y) - H(Y|X) = \Omega(1/3) - \Omega(1/3) = 0 \text{ bits/symb.}$$

$$I(X,Y) = H(X) - H(X|Y) = \log_2 6 - \log_2 6 = 0 \text{ bits/symb.}$$

$$I(X,Y) = H(X) + H(Y) - H(X,Y) = \log_2 6 + \Omega(1/3) - 3.503 = 0 \text{ bits/symb.}$$

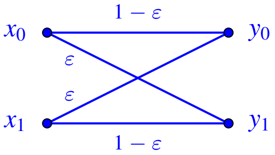
$$I(X,Y) = 0 \text{ bits/symb. (INDEPENDENT!!!)}$$

Examples

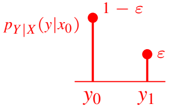
of

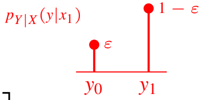
Channel Capacity

Binary symmetric channel with $BER = \varepsilon$



- Capacity: $C = \max_{p_X(x_i)} I(X,Y)$
- Unknowns: $p_X(x_0)$, $p_X(x_1)$
- Evaluation of $I(X,Y)$





$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^{1} p_X(x_i)\; H(Y|X=x_i)$$

$$= H(Y) - \sum_{i=0}^{1} p_X(x_i) \underbrace{\left[-\sum_{j=0}^{1} p_{Y|X}(y_j|x_i)\;\log p_{Y|X}(y_j|x_i)\right]}_{-\varepsilon\;\log(\varepsilon)-(1-\varepsilon)\;\log(1-\varepsilon)=\Omega(\varepsilon)} = H(Y) - \Omega(\varepsilon)$$

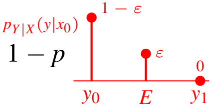
- Channel capacity
  - Maximum of the mutual information
    - ★ $H(Y|X) = \Omega(\varepsilon)$ does not depend on $p_X(x_i)$ for this channel
      - $I(X,Y)$ is maximum when $H(Y)$ is maximum
    - ★ $H(Y)$ is maximum if $p_Y(y_j) = \frac{1}{M_Y}$: $\max H(Y) = \log M_Y = 1$ bit/símb.

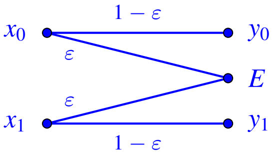
$$\left.\begin{array}{l} p_Y(y_0) = \frac{1}{2} = p_X(x_0)(1-\varepsilon) + p_X(x_1)\;\varepsilon \\ p_Y(y_1) = \frac{1}{2} = p_X(x_0)\;\varepsilon + p_X(x_1)(1-\varepsilon) \end{array}\right\} \rightarrow p_X(x_0) = \frac{1}{2},\; p_X(x_1) = \frac{1}{2}$$

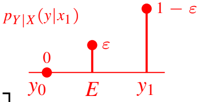
$$\boxed{C = 1 - \Omega(\varepsilon) \text{ bits/use}} \quad \boxed{p_X(x_0) = p_X(x_1) = \frac{1}{2}}$$

Binary Erasure Channel (BEC)

- Capacty: $C = \max_{p_X(x_i)} I(X,Y)$
- Unknowns: $p_X(x_0) = p$, $p_X(x_1) = 1-p$
- Evaluation of $I(X,Y)$







$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^{1} p_X(x_i)\; H(Y|X=x_i)$$

$$= H(Y) - \sum_{i=0}^{1} p_X(x_i) \underbrace{\left[-\sum_{j=0}^{1} p_{Y|X}(y_j|x_i)\; \log p_{Y|X}(y_j|x_i) \right]}_{-\varepsilon\ \log(\varepsilon) - (1-\varepsilon)\ \log(1-\varepsilon) = \Omega(\varepsilon)} = H(Y) - \Omega(\varepsilon)$$

- Channel capacity
  - Maximum of the mutual information
    - ★ $H(Y|X) = \Omega(\varepsilon)$ does not depends on $p_X(x_i)$ for this channel
      - $I(X,Y)$ is maximum when $H(Y)$ is maximum
    - ★ Probabilities $p_Y(y_j)$ to compute $H(Y)$

$$\begin{cases} p_Y(y_0) = p(1-\varepsilon) \\ p_Y(E) = p\varepsilon + (1-p)\varepsilon = \varepsilon \\ p_Y(y_1) = (1-p)(1-\varepsilon) \end{cases}$$

$$H(Y) = -p(1-\varepsilon)\log[p(1-\varepsilon)] - (1-p)(1-\varepsilon)\log[(1-p)(1-\varepsilon)] - \varepsilon\log\varepsilon$$

Binary Erasure Channel (BEC)

$$\begin{aligned}
H(Y) &= -p(1-\varepsilon)\log[p(1-\varepsilon)] \\
&\quad -(1-p)(1-\varepsilon)\log[(1-p)(1-\varepsilon)] \\
&\quad -\varepsilon\log\varepsilon \\
&= -p(1-\varepsilon)[\log p + \log(1-\varepsilon)] \\
&\quad -(1-p)(1-\varepsilon)[\log(1-p) + \log(1-\varepsilon)] \\
&\quad -\varepsilon\log\varepsilon \\
&= -p(1-\varepsilon)\log p - (1-p)(1-\varepsilon)\log(1-p) \\
&\quad -\underbrace{[p(1-\varepsilon) + (1-p)(1-\varepsilon)]}_{(1-\varepsilon)}\log(1-\varepsilon) \\
&\quad -\varepsilon\log\varepsilon \\
&= (1-\varepsilon)\Omega(p) + \Omega(\varepsilon)
\end{aligned}$$

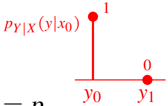
$$I(X,Y) = H(Y) - \Omega(\varepsilon) = (1-\varepsilon)\Omega(p)$$

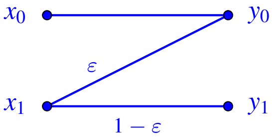
$C = 1-\varepsilon$ bits/use

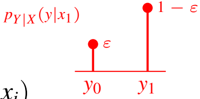
$p_X(x_0) = p = \frac{1}{2},\ p_X(x_1) = 1-p = \frac{1}{2}$

Z channel with error prob. ε

- Capacity: $C = \max_{p_X(x_i)} I(X,Y)$
- Unknowns: $p_X(x_0) = 1-p$, $p_X(x_1) = p$
- Evaluation of $I(X,Y)$







$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^{1} p_X(x_i)\, H(Y|X=x_i)$$

$$H(Y|X=x_0) = 0, \qquad H(Y|X=x_1) = \Omega(\varepsilon) \text{ bits/simb.}$$

$$= H(Y) - p\Omega(\varepsilon)$$

$$\left\{ \begin{array}{l} p_Y(y_0) = 1-p+p\varepsilon = 1-p(1-\varepsilon) \\ p_Y(y_1) = p(1-\varepsilon) \end{array} \right\} \rightarrow H(Y) = \Omega(p(1-\varepsilon))$$

$$I(X,Y) = H(Y) - H(Y|X) = \Omega(p(1-\varepsilon)) - p\Omega(\varepsilon)$$
$$= -p(1-\varepsilon)\log[p(1-\varepsilon)] - (1-p(1-\varepsilon))\log[1-p(1-\varepsilon)] - p\Omega(\varepsilon)$$

Z channel with error prob. ε

$$\begin{aligned} I(X,Y) &= H(Y) - H(Y|X) = \Omega(p(1-\varepsilon)) - p\Omega(\varepsilon) \\ &= -p(1-\varepsilon)\log[p(1-\varepsilon)] - (1-p(1-\varepsilon))\log[1-p(1-\varepsilon)] - p\Omega(\varepsilon) \end{aligned}$$

$$\boxed{\Omega(x) = -x\log x - (1-x)\log(1-x)} \rightarrow \boxed{\frac{\partial\Omega(x)}{\partial x} = \log\frac{1-x}{x}}$$

$$\frac{\partial I(X,Y)}{\partial p} = (1-\varepsilon)\log\frac{1-p(1-\varepsilon)}{p(1-\varepsilon)} - \Omega(\varepsilon) = 0$$

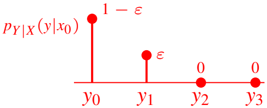
$$\log\frac{1-p(1-\varepsilon)}{p(1-\varepsilon)} = \frac{\Omega(\varepsilon)}{1-\varepsilon} \rightarrow \frac{1-p(1-\varepsilon)}{p(1-\varepsilon)} = 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}} \rightarrow p = \frac{1}{(1-\varepsilon)\left(1+2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}$$

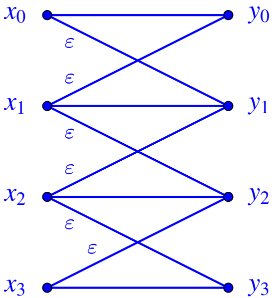
$$\boxed{C = \Omega\left(\frac{1}{1+2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}}\right) - \frac{1}{(1-\varepsilon)\left(1+2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}\Omega(\varepsilon) \text{ bits/use}}$$

$$\boxed{p_X(x_0) = 1-p = 1-\frac{1}{(1-\varepsilon)\left(1+2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}, \qquad p_X(x_1) = p = \frac{1}{(1-\varepsilon)\left(1+2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}}$$

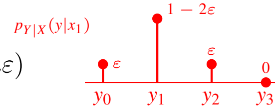
Other channels (symmetric)

- Capacity: $C = \max_{p_X(x_i)} I(X,Y)$
- Unknowns: $p_i = p_X(x_i)$, $i \in \{0,1,2,3\}$
- Evaluation of $I(X,Y) = H(Y) - H(Y|X)$





$$H(Y|X=x_0) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon) = \Omega(\varepsilon)$$

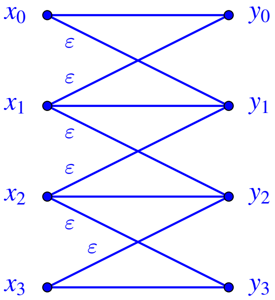


$$\begin{aligned} H(Y|X=x_1) &= -2 \times \varepsilon \log \varepsilon - (1-2\varepsilon)\log(1-2\varepsilon) \\ &= \Omega(2\varepsilon) + 2\varepsilon \end{aligned}$$

$$H(Y|X=x_2) = H(Y|X=x_1), \qquad H(Y|X=x_3) = H(Y|X=x_0)$$

$$H(Y|X) = (p_0+p_3)\Omega(\varepsilon) + (p_1+p_2)\left[\Omega(2\varepsilon) + 2\varepsilon\right]$$

Other channels (symmetric)



$$p_X(x_0) = p_0, p_X(x_1) = p_1, p_X(x_2) = p_2, p_X(x_3) = p_3$$

$$\left\{ \begin{array}{l} p_Y(y_0) = p_0(1-\varepsilon) + p_1\varepsilon \\ p_Y(y_1) = p_1(1-2\varepsilon) + (p_0+p_2)\varepsilon \\ p_Y(y_2) = p_2(1-2\varepsilon) + (p_1+p_3)\varepsilon \\ p_Y(y_3) = p_3(1-\varepsilon) + p_2\varepsilon \end{array} \right\}$$

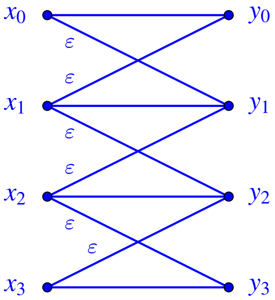
$$\begin{aligned} H(Y) = & -(p_0(1-\varepsilon)+p_1\varepsilon)\log(p_0(1-\varepsilon)+p_1\varepsilon) \\ & -(p_1(1-2\varepsilon)+(p_0+p_2)\varepsilon)\log(p_1(1-2\varepsilon)+(p_0+p_2)\varepsilon) \\ & -(p_2(1-2\varepsilon)+(p_1+p_3)\varepsilon)\log(p_2(1-2\varepsilon)+(p_1+p_3)\varepsilon) \\ & -(p_3(1-\varepsilon)+p_2\varepsilon)\log(p_3(1-\varepsilon)+p_2\varepsilon) \end{aligned}$$

$$H(Y|X) = (p_0+p_3)\Omega(\varepsilon) + (p_1+p_2)\left[\Omega(2\varepsilon)+2\varepsilon\right]$$

$$\boxed{I(X,Y) = H(Y) - H(Y|X)}$$

Other channels (symmetric) - Symmetry

$$p_X(x_0) = p_3, p_X(x_1) = p_2, p_X(x_2) = p_1, p_X(x_3) = p_0$$



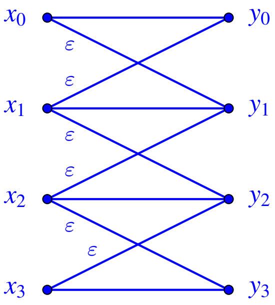
$$\left\{ \begin{array}{l} p_Y(y_0) = p_3(1-\varepsilon) + p_2\varepsilon \\ p_Y(y_1) = p_2(1-2\varepsilon) + (p_3+p_1)\varepsilon \\ p_Y(y_2) = p_1(1-2\varepsilon) + (p_2+p_0)\varepsilon \\ p_Y(y_3) = p_0(1-\varepsilon) + p_1\varepsilon \end{array} \right\}$$

$$\begin{aligned} H(Y) = & -(p_0(1-\varepsilon)+p_1\varepsilon)\log(p_0(1-\varepsilon)+p_1\varepsilon) \\ & -(p_1(1-2\varepsilon)+(p_0+p_2)\varepsilon)\log(p_1(1-2\varepsilon)+(p_0+p_2)\varepsilon) \\ & -(p_2(1-2\varepsilon)+(p_1+p_3)\varepsilon)\log(p_2(1-2\varepsilon)+(p_1+p_3)\varepsilon) \\ & -(p_3(1-\varepsilon)+p_2\varepsilon)\log(p_3(1-\varepsilon)+p_2\varepsilon) \end{aligned}$$

$$H(Y|X) = (p_0+p_3)\Omega(\varepsilon) + (p_1+p_2)\left[\Omega(2\varepsilon)+2\varepsilon\right]$$

$$\boxed{I(X,Y) = H(Y) - H(Y|X)}$$

Other channels (symmetric) - Parameterization



$$p_X(x_0) = p_X(x_3) = \frac{p}{2}, \quad p_X(x_1) = p_X(x_2) = \frac{1-p}{2}$$

$$\left\{ \begin{array}{l} p_Y(y_0) = \frac{p}{2}(1-\varepsilon) + \frac{1-p}{2}\varepsilon = \frac{p+\varepsilon-2p\varepsilon}{2} \\ p_Y(y_1) = \frac{1-p}{2}(1-2\varepsilon) + \frac{1}{2}\varepsilon = \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \\ p_Y(y_2) = \frac{1-p}{2}(1-2\varepsilon) + \frac{1}{2}\varepsilon = \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \\ p_Y(y_3) = \frac{p}{2}(1-\varepsilon) + \frac{1-p}{2}\varepsilon = \frac{p+\varepsilon-2p\varepsilon}{2} \end{array} \right\}$$

$$\begin{aligned} H(Y) = & -2 \times \frac{p+\varepsilon-2p\varepsilon}{2} \log \frac{p+\varepsilon-2p\varepsilon}{2} \\ & -2 \times \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \log \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \\ & = -(p+\varepsilon-2p\varepsilon)[\log(p+\varepsilon-2p\varepsilon) - \log 2] \\ & \quad -(1-(p+\varepsilon-2p\varepsilon))[\log(1-(p+\varepsilon-2p\varepsilon)) - \log 2] \\ & = 1 + \Omega(p+\varepsilon-2p\varepsilon) \end{aligned}$$

$$H(Y|X) = p\Omega(\varepsilon) + (1-p)\left[\Omega(2\varepsilon) + 2\varepsilon\right]$$

$$I(X,Y) = 1 + \Omega(p+\varepsilon-2p\varepsilon) - p\Omega(\varepsilon) - (1-p)\left[\Omega(2\varepsilon) + 2\varepsilon\right]$$

Other channels (symmetric) - Capacity

$$I(X,Y) = 1 + \Omega(p+\varepsilon-2p\varepsilon) - p\Omega(\varepsilon) - (1-p)\left[\Omega(2\varepsilon)+2\varepsilon\right]$$

$$\frac{\partial I(X,Y)}{\partial p} = (1-2\varepsilon)\log\frac{1-(p+\varepsilon-2p\varepsilon)}{p+\varepsilon-2p\varepsilon} - \Omega(\varepsilon) + \Omega(2\varepsilon) + 2\varepsilon = 0$$

$$\log\frac{1-(p+\varepsilon-2p\varepsilon)}{p+\varepsilon-2p\varepsilon} = \frac{\Omega(\varepsilon)-\Omega(2\varepsilon)-2\varepsilon}{1-2\varepsilon} \rightarrow \frac{1-(p+\varepsilon-2p\varepsilon)}{p+\varepsilon-2p\varepsilon} = 2^{\frac{\Omega(\varepsilon)-\Omega(2\varepsilon)-2\varepsilon}{1-2\varepsilon}}$$

$$p = \frac{1-\varepsilon\left(1+2^{\frac{\Omega(\varepsilon)-\Omega(2\varepsilon)-2\varepsilon}{1-2\varepsilon}}\right)}{(1-2\varepsilon)\left(1+2^{\frac{\Omega(\varepsilon)-\Omega(2\varepsilon)-2\varepsilon}{1-2\varepsilon}}\right)}$$

$$C = 1 + \Omega(p+\varepsilon-2p\varepsilon) - p\Omega(\varepsilon) - (1-p)\left[\Omega(2\varepsilon)+2\varepsilon\right] \text{ bits/use}$$

$$p_X(x_0) = p_X(x_3) = \frac{p}{2}, \quad p_X(x_1) = p_X(x_2) = \frac{1-p}{2}$$