Joint entropy

A joint ensemble (X, Y) with probability

$$p(x, y) = \text{Prob}(X = x, Y = y)$$

Marginals

$$p(x) = \sum_y p(x, y), \quad p(y) = \sum_x p(x, y)$$

Conditionals

$$p(x|y) = p(x, y)/p(y), \quad p(y|x) = p(x, y)/p(x)$$

Joint entropy

$$H(X, Y) = \sum_{x,y} -p(x, y) \log p(x, y)$$

Marginal entropies

$$H(X) = \sum_{x,y} -p(x, y) \log \left(\sum_y p(x, y) \right)$$

$$H(Y) = \sum_{x,y} -p(x, y) \log \left(\sum_x p(x, y) \right)$$

Conditional entropy

$$H(X|Y) = \sum_{x,y} -p(x,y) \log p(x|y)$$

$$H(Y|X) = \sum_{x,y} -p(x,y) \log p(y|x)$$

An identity:

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y) \end{aligned}$$

Interpretation: $H(X|Y)$ is the number of bits to specify X when Y is common knowledge.

Conditioning cannot increase entropy

$$
\begin{aligned}
& H(X|Y) - H(X) \\
&\quad = \sum_{x,y} p(x,y) \log \frac{p(x)}{p(x|y)} \\
&\quad \leq \log(e) \sum_{x,y} p(x,y) \left(\frac{p(x)}{p(x|y)} - 1 \right) \\
&\quad = \log(e) \left[\sum_{x,y} p(x)p(y) - \sum_{x,y} p(x,y) \right] \\
&\quad = 0,
\end{aligned}
$$

with equality iff $p(x,y) = p(x)p(y)$, i.e., X and Y are independent.

Mutual information

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

Symmetry: $I(Y;X) = I(X;Y)$

Nonnegativity: $I(X;Y) \geq 0$, with equality iff X and Y are independent.

Chain rule

$$
\begin{aligned}
& H(X_1, X_2, X_3) \\
& \quad = \; H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2)
\end{aligned}
$$

$$H(X_1, \ldots, X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_1, \ldots, X_{i-1})$$

Markov chains

$X_1, \ldots, X_n$ is a Markov chain if

$$p(x_1, \ldots, x_n) = p(x_1) \prod_{i=2}^{n} p(x_i | x_{i-1}).$$

Notation: $X_1 \to X_2 \to \cdots \to X_n$

Strong subadditivity

$$H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z)$$

with equality iff $X \to Y \to Z$.

Data processing

Let $X \to Y \to Z$. Then,

$$I(X;Z) \leq I(X;Y)$$

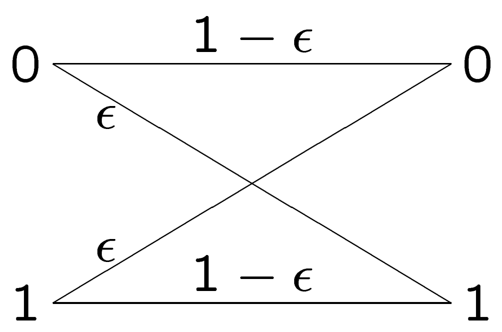
Processing Y can only destroy information that Y carries about X.

Communication channels

An input alphabet: $\mathcal{X}$
An output alphabet: $\mathcal{Y}$
Transition probabilities:

$$p(y|x) = \text{Prob}(y \text{ received } | \ x \text{ sent})$$

Example: Binary symmetric channel (BSC)



ϵ = the crossover probability

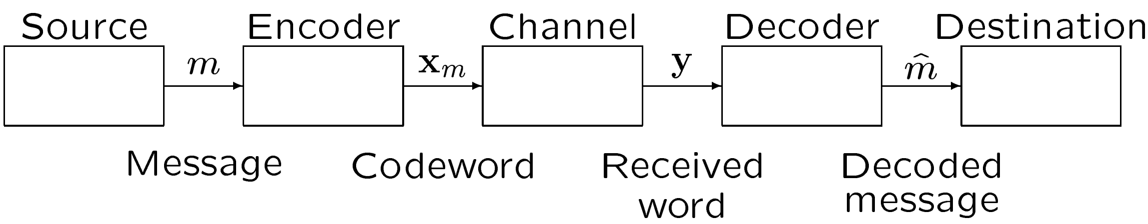
Memoryless channels

Input sequence $\mathbf{x} = (x_1, \ldots, x_N)$ sent,
output sequence $\mathbf{y} = (y_1, \ldots, y_N)$ received
with probability

$$p(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} p(y_n|x_n)$$

Independent transitions = Memoryless channel

Communication systems



Message: $m \in \{1, \ldots, M\}$

Encoder: $m \to \mathbf{x}_m = (x_{m,1}, \ldots, x_{m,N})$

Channel: $\mathbf{x}_m \to \mathbf{y}$ with probability

$$p(\mathbf{y}|\mathbf{x}_m) = \prod_{n=1}^{N} p(y_n|x_{m,n})$$

Decoder: $\mathbf{y} \to \widehat{m} \in \{1, \ldots, M\}$

Design goals

Maximize the rate

$$R = \frac{1}{N} \log_2 M \ \text{ bits}$$

Minimize the probability of decoding error

$$P_e = \text{Prob}[\widehat{m} \neq m]$$

Minimize encoder-decoder complexity

Channel capacity

$$C = \max_{p(x)} \; I(X;Y)$$
$$= \max_{p(x)} \sum_{x,y} p(x)p(y|x) \log \left(\frac{p(y|x)}{\sum_{x'} p(x')p(y|x')} \right)$$

Maximization over all probability assignments on channel input alphabet.

Example: BSC capacity

$C = 1 - H(\epsilon) = 1 + \epsilon \log(\epsilon) + (1 - \epsilon) \log(1 - \epsilon)$

achieved with $p(0) = p(1) = 1/2$

Noisy channel coding theorem

Consider a DMC with capacity C.

Achievability: For any given $R < C$ and $\delta > 0$, there exists an encoder-decoder pair with rate $\geq R$ and $P_e \leq \delta$.

Converse: For $R > C$, there exists a $\delta > 0$ such that for any encoder-decoder pair with rate $\geq R$, one has $P_e \geq \delta$.