Lecture 2: Some Information Theory

see: Classical Information Theory &
Classical Error Correction

in: Lecture Notes on Quantum Information
D. Bruß & G. Leuchs (see webpage)

Def. Shannon entropy

Let S be a source emitting symbols $x_1, \ldots, x_n$ with probabilities $p(x_i) = p_i$, with identical independent distribution (iid).
Then the Shannon entropy of S is given by

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

Independence $p(x_i x_j) = p(x_i)\, p(x_j)$

$$H(S \times S) = -\sum_{i,j=1}^{n} p(x_i x_j) \log_2 p(x_i x_j)$$

$$= -\sum_{i,j} p_i p_j \log_2 (p_i p_j)$$

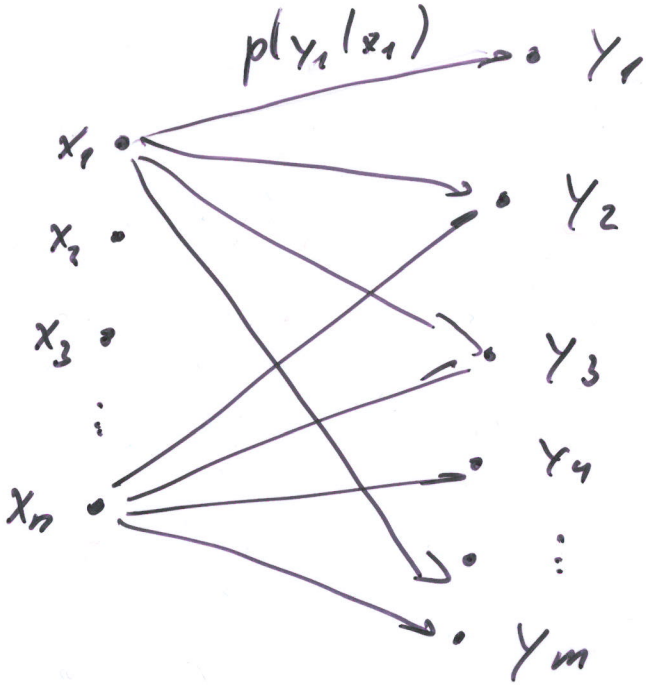
$$= -\sum_{i,j} p_i p_j \log_2 p_i - \sum_{i,j} p_i p_j \log_2 p_j$$

$$= -\sum_{i} p_i \log_2 p_i - \sum_{j} p_j \log_2 p_j$$

$$= 2\, H(S)$$

recall: channel with input alphabet $X = \{x_1, \dots, x_n\}$ and output alphabet $Y = \{y_1, \dots, y_m\}$

discrete memoryless channel described by transition probabilities $p(y_j \mid x_i)$



edges are labelled by non-zero transition prob.

Def. joint entropy of the input X and output Y

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y)$$

Consider input and output as one symbol

Def.: conditional entropy

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x)\, H(Y \mid X = x)$$

$$= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log_2 p(y|x)$$

$$= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 p(y|x)$$

$H(Y|X)$ is a measure for the information we additionally get when considering X and Y together, w.r.t. only X.

Fact: chain rule

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

Def.: mutual information

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

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

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

Channel capacity:

How much information can be sent reliably through a channel?

For a discrete memoryless channel, the capacity is:

$$C := \max_{p(x)} I(X;Y) = \max_{p(x)} \left(H(X) - H(X|Y) \right)$$

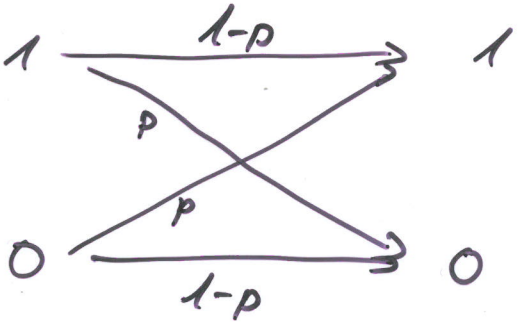
where the maximization is over all input distributions. "encodings"

Shannon's 2nd fundamental coding theorem

noisy coding theorem

Let S be a source with entropy $H(S)$ and let a discrete memoryless channel have capacity C.
If $H(S) < C$ then there exists an encoding scheme such that the output of the source can be transmitted over the channel with an arbitrary small probability of error. Not possible for $H(S) > C$

Example Capacity of the BSC



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

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x p(x) H(Y|X=x) \\ &= H(Y) - \sum_x p(x) H(p) \\ &= H(Y) - H(p) \\ &\leq 1 - H(p) \end{aligned}$$

binary entropy function

$$H(p) := -p \log_2 p - (1-p)\log_2(1-p)$$