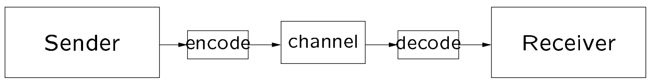
Information Theory: Motivation

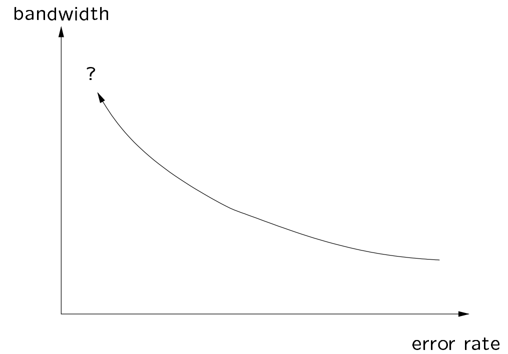
Task: sending data over a (possibly noisy) channel



Problem: if the channel is unreliable because of noise, how much "effort" (bandwidth) is needed to ensure a transmission with an error rate less than ϵ?

Quantification of Data Transmission

Question: how will the required bandwidth grow for an arbitrarily small error rate ϵ? Would it grow to infinity?



Shannon's Central Result (1948): for a channel with given noise, to transmit data with an error rate $\epsilon \to 0$, only a finite bandwidth is needed.

Universal Theory: connection with physics (Jaynes 1957)

Foundations

Preliminaries:

- consider random variables $X, Y, Z \ldots$. Their values will be denoted by $x, y, z \ldots$.
- write probabilities as $P(X = x, Y = y)$ or equivalently $p(x, y)$

Entropy: let X be a random variable assuming values in $\{x_1, x_2, \ldots, x_n\}$. Then define the *entropy* of X as

$$H(X) := -\sum_i p(x_i) \log p(x_i)$$

Remark: if 'log' is a binary logarithm, measure $H(X)$ in *bit*. Assuming a symbol source generating symbols x_i with the probability x_i. The entropy $H(X)$ in bit measures how many symbols $\{0, 1\}$ are required *on average* using optimal coding to send signals from this source over a noiseless channel.

Further Notions

Mutual Information: let X, Y be random variables with joint distribution $p(x, y)$. Then the mutual information $I(X; Y)$ is given by

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

with $H(Y|X) = \sum_x p(x) \cdot H(Y|X = x)$. Here, $H(X, Y)$ is the entropy of the joint random variable (X, Y) and $H(X)$ is the entropy of the marginal distribution for (X, Y) (i.e. $p(x) = \sum_y p(x, y)$), analogously for Y.

Theorems: one has the relations

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

The Kullback-Leibler Distance

Kullback-Leibler Distance: for a random variable with two different distributions p and q, define the Kullback-Leibler distance between p and q as

$$D_{p||q} = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Theorem:

1. $D_{p||q} \geq 0$ and $D_{p||q} \Leftrightarrow p = q$
2. Let $p(x,y)$ be a the joint distribution of random variables X and Y. Write $\tilde{p}(x,y) = p(x)p(y)$ with $p(x)$ and $p(y)$ the marginal distributions of $p(x,y)$. Then

$$I(X;Y) = D_{p||\tilde{p}}$$

Additional Remarks

Remark: if X, Y, Z are continuous, replace $\sum$ by $\int$. However, then the corresponding quantities of H can not be interpreted as an uncertainty, but only quantities like I, D etc. that are differences of entropies and, hence, informational quantities. In that case, write h instead of H for clarification.

Remark: The mutual information $I(X;Y)$ quantifies how much X tells about Y $\Rightarrow$ thus it serves as a measure of information transmission.