Information Channels

Suppose that data must be sent through a *channel* before it can be used. This channel may be unreliable, but we can use a *code* designed counteract this.

Some questions we aim to answer:

- Can we quantify how much information a channel can transmit?
- If we have low tolerance for errors, will we be able to make full use of a channel, or must some of the channel's capacity be lost to ensure a low error probability?
- How can we correct (or at least detect) errors in practice?
- Can we do as well in practice as the theory says is possible?

Error Correction for Memory Blocks

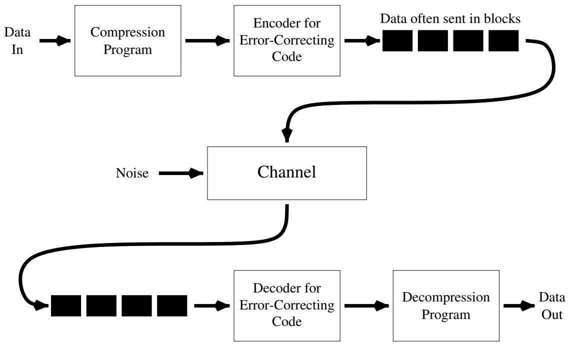
The "channel" may transmit information through time rather than space — ie, it is a memory device.

Many memory devices store data in blocks — eg, 64 bits for RAM, 512 bytes for disk.

Can we correct some errors by adding a few more bits? For instance, could we correct any single error if we use 71 bits to encode a 64 bit block of data stored in RAM?

Error Correction in a Communications System

In other applications, data arrives in a continuous stream. An overall system might look like this:



Error Detection

We might also be interested in detecting errors, even if we can't correct them:

- For RAM or disk memory, error detection tells us that we need to call the repair person.
- For some communication applications, we have the option of asking the sender to re-transmit.
- If we know that a bit is in error, we can try to minimize the damage — eg, if the bit is part of a pixel in an image, we can replace the pixel with the average of nearby pixels.

Formal Definition of a Channel

A channel is defined by

- An input alphabet, $\mathcal{A}_X$, with symbols $a_1, \ldots, a_r$. We will usually assume that the input alphabet is binary, with $\mathcal{A}_X = \{0, 1\}$.

- An output alphabet, $\mathcal{A}_Y$, with symbols called $b_1, \ldots, b_s$. This is also often binary, with $\mathcal{A}_Y = \{0, 1\}$, but in general it can be different from $\mathcal{A}_X$.

- A description of how the output depends on the input.

Channel Transition probabilities

We will assume that the correspondence of input symbols with output symbols is always known — there are no "insertions" or "deletions" of symbols.

We will also assume that the channel is *memoryless* — each output symbol is influenced only by the corresponding input symbol, not by earlier input or output symbols.

The behaviour of such a channel is defined by its *transition probabilities*:

$$Q_{j|i} \;=\; P(Y = b_j \,|\, X = a_i)$$

These transition probabilities are fixed by the nature of the channel. They can be changed only by redesigning it.

The Binary Symmetric Channel (BSC)

For the BSC, the input and output alphabets are both $\{0, 1\}$.

With probability f, the symbol received is different from the symbol transmitted. With probability $1-f$, the symbol is received correctly.

We can view the input and output alphabets as Z_2, the field of integers modulo 2. The channel can then be seen as adding "noise" to the input:

$$b \;=\; a + n$$

where n is 0 with probability $1-f$ and 1 with probability f.

Addition modulo 2 is the same as exclusive-or, and works as follows:

$$0+0 \;=\; 0, \;\; 0+1 \;=\; 1, \;\; 1+0 \;=\; 1, \;\; 1+1 \;=\; 0$$

The Binary Erasure Channel (BEC)

For the BEC, the input alphabet is $\{0, 1\}$, but the output alphabet is $\{0, ?, 1\}$. The "?" output represents an "erasure", in which the transmitted symbol is lost.

An erasure happens with probability f; otherwise, the symbol is received correctly.

The transition probabilities for the BEC can be arranged in a matrix as follows:

$$\boldsymbol{Q} \;=\; (Q_{j|i}) \;=\; \begin{bmatrix} 1-f & 0 \\ f & f \\ 0 & 1-f \end{bmatrix}$$

The Z Channel

The Z Channel has input alphabet is $\{0, 1\}$, and output alphabet $\{0, 1\}$, like the BSC.

However, the Z channel is asymmetrical. The 0 symbol is always transmitted correctly, but the 1 symbol is received incorrectly (as 0) with probability f.

The matrix of transition probabilities for the Z channel is as follows:

$$\boldsymbol{Q} \;=\; (Q_{j|i}) \;=\; \begin{bmatrix} 1 & f \\ 0 & 1-f \end{bmatrix}$$

Channel Input Distribution

We can choose what input symbols we feed into the channel. We might send symbols from some source, the output of a data compression program applied to that source, or an error-correcting code for either of these.

For the moment, we'll assume that the symbols we put in are independent of each other, with some specified distribution:

$$p_i \;=\; P(X = a_i)$$

We might aim to choose these *input probabilities* so that we make efficient use of the channel.

Deriving Some More Probabilities

The input and the transition probabilities together define the *joint probability* for any combination of channel input and output:

$$\begin{aligned} R_{ij} &= P(X = a_i,\, Y = b_j) \\ &= P(X = a_i)\, P(Y = b_j \,|\, X = a_i) \\ &= p_i\, Q_{j|i} \end{aligned}$$

We can now find the *output probabilities*:

$$q_j \;=\; P(Y = b_j) \;=\; \sum_{i=1}^{r} R_{ij} \;=\; \sum_{i=1}^{r} p_i\, Q_{j|i}$$

Finally, we get the *backward probabilities*:

$$\begin{aligned} P(X = a_i \,|\, Y = b_j) &= P(X = a_i,\, Y = b_j) / P(Y = b_j) \\ &= R_{ij}/q_j \;=\; S_{i|j} \end{aligned}$$

The backward probabilities give the situation from the receiver's point of view — given that I've received symbol b_j, how likely is it that the symbol sent was a_i?

Input, Output, and Joint Entropies

The amount of information being sent can be measured by the *input entropy*:

$$H(X) \;=\; \sum_{i=1}^{r} p_i \log(1/p_i)$$

where $p_i = P(X = a_i)$.

Similarly, the amount of "information" received (some of which may actually be noise) is measured by the *output entropy*:

$$H(Y) \;=\; \sum_{j=1}^{s} q_j \log(1/q_j)$$

where $q_j = P(Y = b_j)$.

We also have the *joint entropy*:

$$H(X, Y) \;=\; \sum_{i=1}^{r} \sum_{j=1}^{s} R_{ij} \log(1/R_{ij})$$

where $R_{ij} = P(X = a_i,\, Y = b_j)$. This is the information obtained by an outside observer who sees both the input and the output.

Mutual Information

We can now define the *mutual information* between the input and the output:

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

The mutual information is meant to represent the amount of information that is being communicated from the sender to the receiver.

This makes intuitive sense: The difference of $H(X) + H(Y)$ and $H(X,Y)$ is the "overlap" in the knowledge of the sender and receiver — due to information having been transmitted.

But the real test of this definition is whether it leads to useful theorems and insights.