Definition: Channel Capacity

We just discovered that for some noisy channels, zero-error communication is very hard, or even impossible. For example, if Alice and Bob have to communicate over a binary symmetric channel (BSC) that has non-zero bit-flip probability, they cannot hope to do any zero-error communication, because the Shannon capacity of the BSC's confusability graph is zero.

We also saw that error-correcting codes can help deal with such inherently noisy channels. Even though the communication error may not become zero, an error-correcting code can increase the probability of receiving the correct message. It does come at a cost, however, because the codewords are longer than the original messages, and so the amount of information that is transmitted *per channel use* does not necessarily increase.

In this final part of the module, we explore the limits of how much information can be sent over a channel if a small error is allowed. Central to our study will be the concept of channel capacity. It reflects the maximum amount of information that could *in principle* be communicated with a single use of a channel. In the next module, we will see how well that theoretical limit can be approached with actual error-correcting codes.

Definition: Channel capacity

The channel capacity C of a discrete, memoryless channel $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ is given by

$$C := \max_{P_X} I(X;Y).$$

Remember that using a certain input distribution P_X for a channel $P_{Y|X}$ yields a joint input-output distribution P_{XY} which determines the real quantity $I(X;Y)$ we can optimize over. One can argue that the maximum is attained and therefore the channel capacity is a well-defined quantity.

Important: the channel capacity is often called the Shannon capacity (of a channel). You should not confuse it with the Shannon Capacity of a Graph.

Generally, the Shannon capacity of a channel is not equal to the Shannon capacity of its confusability graph.

Example: Capacity of a BSC

What is the capacity (in terms of f) of a binary symmetric channel with parameter $f \in [0, 1/2]$?

Show hint

Rewrite $I(X;Y)$ as $H(Y) - H(Y|X)$ and note that you can compute $H(Y|X)$ without fixing P_X. Then think about how to maximize $H(Y)$.

Show solution

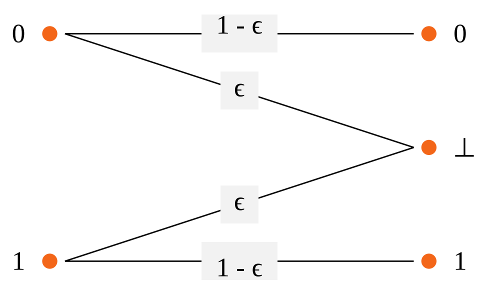
The channel capacity is

$$\begin{aligned}\max_{P_X} I(X;Y) &= \max_{P_X} (H(Y) - H(Y|X)) \\ &= \max_{P_X} \left(H(Y) - \sum_{x \in \mathcal{X}} P_X(x) \cdot H(Y|X=x) \right) \\ &= \max_{P_X} \left(H(Y) - \sum_{x \in \mathcal{X}} P_X(X) \cdot h(f) \right) \\ &= \max_{P_X} (H(Y) - h(f)) \\ &= 1 - h(f).\end{aligned}$$

The last step follows because $H(Y)$ is maximized if Y is uniform, which is achievable by choosing X to be uniform.

Example: Capacity of a BEC

Consider the binary erasure channel (BEC) with $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, \perp\}$, where $\perp$ is the **erasure symbol**, and $\epsilon \in [0, 1]$ is the **erasure probability**:



What is the channel capacity of the BEC, as a function of ϵ?

Show hint

Contrary to the previous example, break $I(X;Y)$ up as $H(X) - H(X|Y)$, using symmetry of the mutual information. Consider the three possible outputs separately: how much uncertainty is left if you receive output 0? What about output 1? And output $\perp$?

Show solution

Write p for $P_X(0)$.

$$\begin{aligned}\max_{P_X} I(X;Y) &= \max_{P_X}\left(H(X) - H(X|Y)\right)\\ &= \max_{p}\left(h(p) - \sum_{y\in\mathcal{Y}} P_Y(y)\cdot H(X|Y=y)\right)\\ &= \max_{p}\left(h(p) - P_Y(\perp)\cdot h(p)\right)\\ &= \max_{p}\left(h(p)(1-\epsilon)\right)\\ &= 1-\epsilon\,.\end{aligned}$$

Again, the last step follows because $H(X) = h(p)$ is maximized if X is uniform, hence $p = \frac{1}{2}$.

If a channel is memoryless, then using it more than once does not increase the capacity *per transmission*. Note that this is different from the zero-error setting, where multiple channel uses can in fact increase the efficiency of getting information across! This is formally captured in the following lemma, which we state without proof:

Lemma: Multiple Channel Uses

Let $X_1,\ldots,X_n =: X^n$ be n random variables. Let Y^n be the result of passing X^n through a discrete memoryless channel of capacity C. Then for any joint distribution P_{X^n},

$$I(X^n, Y^n) \leq n \cdot C.$$