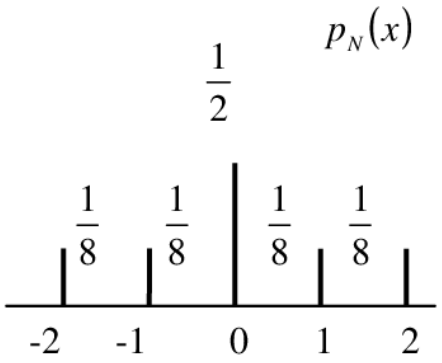
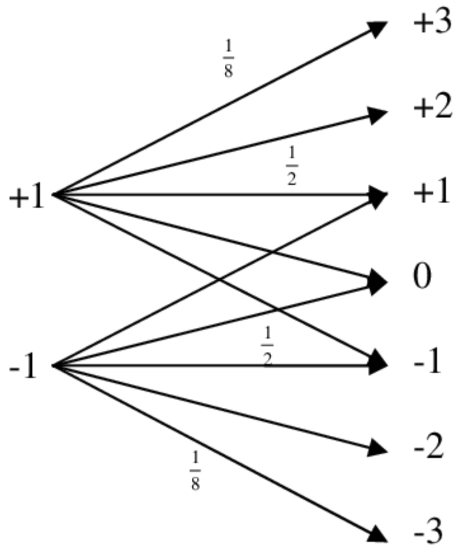
Concatenated Channels & Data Processing Inequality

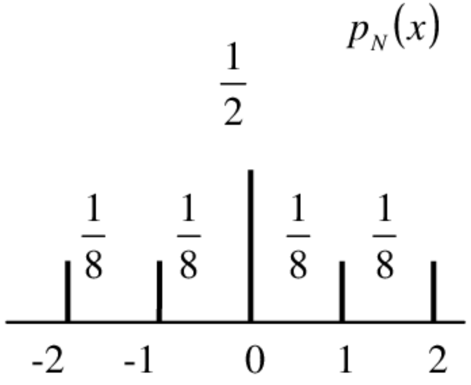
Post-processing on channel outputs

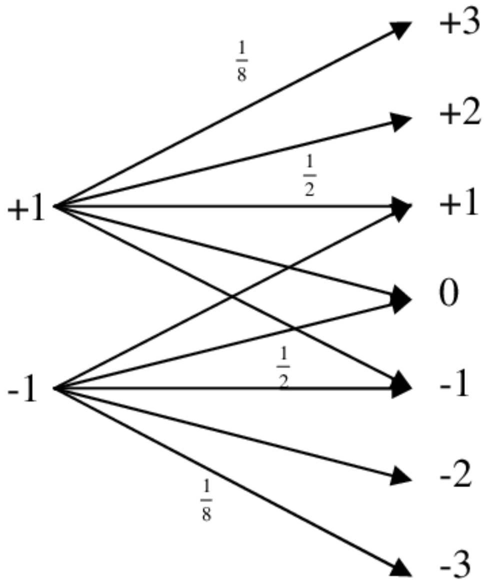
Example

Consider the following channel with binary input alphabet $\mathcal{X} = \{-1, +1\}$ and additive noise N, $\mathcal{N} = \{-2, -1, 0, 1, 2\}$ with corresponding pdf $\left\{\frac{1}{8}, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right\}$.









Capacity computation:

By symmetry the capacity is maximised for $p_X(0) = p_X(1) = \frac{1}{2}$

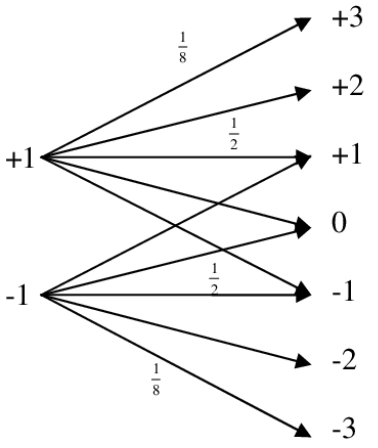
$H(Y|X) = H(N) = H\left(\left\{\frac{1}{8}, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right\}\right) = \frac{1}{2}\log 2 + \frac{4}{8}\log 8 = 2\text{bits}$

now:

$H(Y) = H\left(\left\{\frac{1}{16}, \frac{1}{16}, \frac{5}{16}, \frac{2}{16}, \frac{5}{16}, \frac{1}{16}, \frac{1}{16}\right\}\right) =$
$\frac{4}{16}\log 16 + \frac{2}{16}\log 8 + \frac{10}{16}\log\frac{16}{5} = 2.4238$ bits

hence:

$C = \max I(X;Y) = H(Y) - H(Y|X) = 2.4238 - 2 = 0.4238$ bits per channel use



- The receiver chooses to process the channel outputs, using the following rounding function:

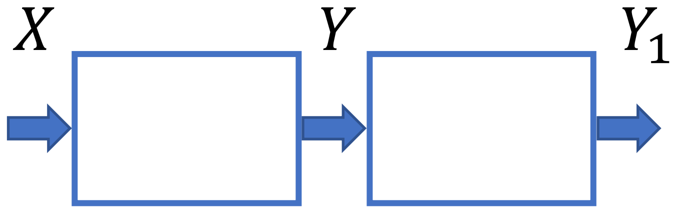
$$Y_1 = \begin{cases} +1 & if \quad y > 0 \\ 0 & if \quad y = 0 \\ -1 & if \quad y < 0 \end{cases}$$

- Now we will recompute the capacity for the channel $p(y_1|x)$. Again by symmetry:

$$C_1 = H(Y_1) - H(Y_1|X) = H\left(\left\{\tfrac{7}{16}, \tfrac{1}{8}, \tfrac{7}{16}\right\}\right) - H\left(\left\{\tfrac{6}{8}, \tfrac{1}{8}, \tfrac{1}{8}\right\}\right) = 1.4186 - 1.0613 = 0.3573\text{bits}$$

- We observe that $C_1 < C$ - i.e. post-processing has reduced the channel capacity.

Let's think about it

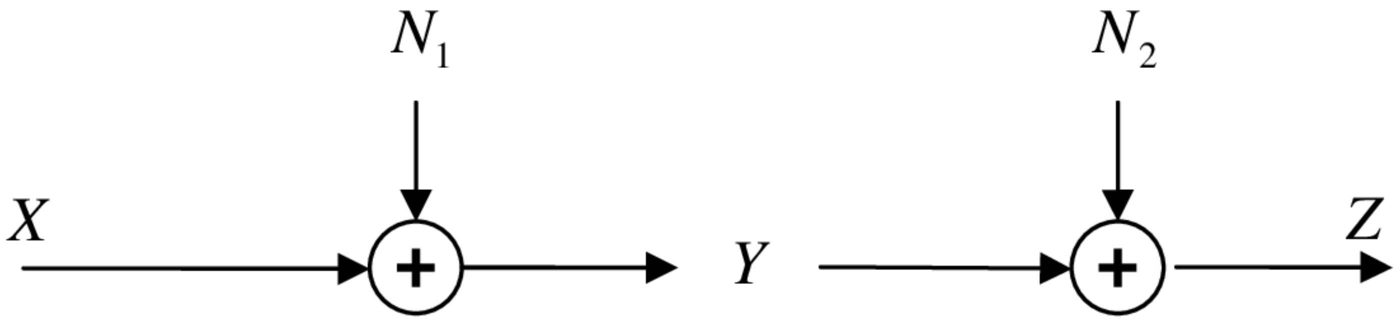


Data processing inequality - intro example

Example

Cascaded binary symmetric channels

Suppose the output Y of BSC #1 is the input to BSC #2 whose output is Z.



This is an example of a Markov chain:

- When random variables X, Y, Z have the property that Z does not depend directly on X but only on a (probabilistic) "function" Y , these random variables are said to form a Markov chain in the order $X \rightarrow Y \rightarrow Z$.

Markov chains

- The defining condition of a Markov chain is:

$$p(z \mid y, x) = p(z \mid y) \Rightarrow p(x, y, z) = p(x)\, p(y \mid x)\, p(z \mid y, x) = p(x)\, p(y \mid x)\, p(z \mid y)$$

- The Markov condition is equivalent to "X and Z are conditionally independent given Y " i.e. $X \perp Z \mid Y$:

$$p(x, z \mid y) = p(x \mid y)\, p(z \mid x, y) = p(x \mid y)\, p(z \mid y)$$

- from which we can deduce:

$$I(X; Z \mid Y = y) = 0 \Rightarrow I(X; Z \mid Y) = 0$$

Conditional mutual information

- Suppose X, Y, Z are random variables with joint pdf $p(x, y, z)$.
- If $p(z) > 0$ then $p(x, y | Z = z)$ is a well defined conditional pdf, and the information theoretic quantities $H(X | Z = z)$,$H(Y | Z = z)$, $H(X, Y | Z = z)$, $H(Y | X, Z = z)$, $I(X; Y | Z = z)$ are all defined in the expected way

Example

$$H(Y | X, Z = z) = \sum_{x,y} p(x, y | z) \log \frac{1}{p(y | x, z)}$$

$$I(X; Y | Z = z) = \sum_{x,y} p(x, y | z) \log \frac{p(x, y | z)}{p(x | z)\, p(y | z)}$$

- We obtain $H(Y|X,Z)$ averaging over Z:

$$H(Y|X,Z) = \sum_{z} p(z) \sum_{x,y} p(x,y|z) \log \frac{1}{p(y|x,z)} =$$
$$\sum_{x,y,z} p(x,y,z) \log \frac{1}{p(y|x,z)}$$

- And $I(X;Y|Z)$ by averaging over Z:

$$I(X;Y|Z) = \sum_{z} p(z) \sum_{x,y} p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} =$$
$$\sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} = \sum_{x,y,z} p(x,y,z) \log \frac{p(y|x,z)}{p(y|z)} =$$
$$\sum_{x,y,z} p(x,y,z) \log \frac{1}{p(y|z)} - \sum_{x,y,z} p(x,y,z) \log \frac{1}{p(y|x,z)} =$$
$$H(Y|Z) - H(Y|X,Z)$$

Conditional mutual information facts

- Conditional mutual information is an average of non-negative mutual informations, so it is non-negative.
- Conditioning reduces entropy but does not in general reduce mutual information!!!

Another useful identity[1]

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

[1](we have used the result from the previous slide $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$)

Data processing inequality

Theorem

If $X \to Y \to Z$ is a Markov chain then $I(X;Y) \geq I(X;Z)$

Proof.

We use a trick - expand $I(X;Y,Z)$ in two ways: $\square$

1) $I(X;Y,Z) = I(X;Y) + I(X;Z|Y) = I(X;Y) + 0 = I(X;Y)$
2) $I(X;Y,Z) = I(X;Z) + I(X;Y|Z) \geq I(X;Z)$ since $I(X;Y|Z) \geq 0$

- The data processing inequality states that no transformations (deterministic or probabilistic) on the raw data can provide more information about the original input.

No signal processing tricks on the channel outputs can increase the capacity

Concatenation of symmetric channels

- Consider the capacity of a chain of L ternary channels, where each channel is symmetric with $p = \{0.9, 0.05, 0.05\}$.