Math 5251 Hamming distance & decoding (§4.4)

If we send codewords $\mathcal{C} \subset \{0,1\}^*$ of length ℓ
$$\mathcal{C} = \{x_1, \dots, x_m\}$$

through a BSC with error probability $p < \frac{1}{2}$ ← safe to assume; Why?

and receive the word y, which word x_i should we decode it as?

You would think we should pick an x_i that minimizes this...

DEF'N: Given two words in Σ^* of same length ℓ

$x = x^{(1)} x^{(2)} \dots x^{(\ell)}$
$y = y^{(1)} y^{(2)} \dots x^{(\ell)}$ their Hamming distance is

$$d(x,y) := \#\{p = 1, 2, \dots, \ell : x^{(p)} \neq y^{(p)}\}$$

EXAMPLES

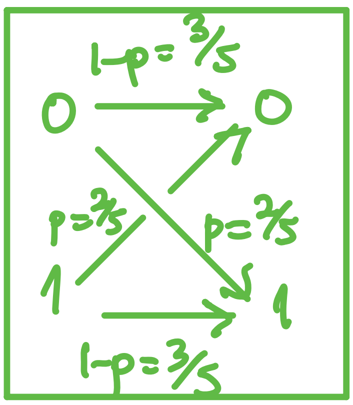
$d(101, 000) = 2,$
$d(102, 000) = 2$
$d(101, 010) = 3$
$d(101, 101) = 0$

If we don't know much about the source word probabilities, this is a good rule to follow and called maximum likelihood estimation, in that it picks x_i maximizing

$$P\left(\underset{\text{received}}{y} \,\middle|\, \underset{\text{sent}}{x_i}\right) = p^{d(x_i,y)}(1-p)^{\ell - d(x_i,y)}$$

EXAMPLE If we send words in $\{0,1\}^*$ of length 2 via repetition code of length $\ell=4$ through a BSC of error prob $p=2/5$

$$C = \begin{array}{ll} x_1 & 00|00 \\ x_2 & 01|01 \\ x_3 & 10|10 \\ x_4 & 11|11 \end{array}$$



receiving $y = 0111$

then how should we decode?

$$P(y=0111 \text{ rec'd} \mid x_1=0000 \text{ sent}) = \left(\tfrac{2}{5}\right)^3\left(\tfrac{3}{5}\right)^1 = \frac{24}{625}$$

$$\boxed{x_2=0101 \qquad = \left(\tfrac{2}{5}\right)^1\left(\tfrac{3}{5}\right)^3 = \frac{54}{625}}$$

$$x_3=1010 \qquad = \left(\tfrac{2}{5}\right)^3\left(\tfrac{3}{5}\right)^1 = \frac{24}{625}$$

$$\boxed{x_4=1111 \qquad = \left(\tfrac{2}{5}\right)^1\left(\tfrac{3}{5}\right)^3 = \frac{54}{625}}$$

decode y as one of these two

What would be an alternative? If we had more info about source probabilities $p_1, \ldots, p_m$, then one could use ideal observer/minimum error rule, maximizing $P(\underset{\text{sent}}{x_i} \mid \underset{\text{received}}{y})$ via a Bayesian calculation:

$$P\left(\underset{\text{sent}}{x_i} \,\middle|\, \underset{\text{received}}{y}\right) = \frac{P\left(\underset{\text{sent}}{x_i} \cap \underset{\text{received}}{y}\right)}{P\left(\underset{\text{received}}{y}\right)}$$

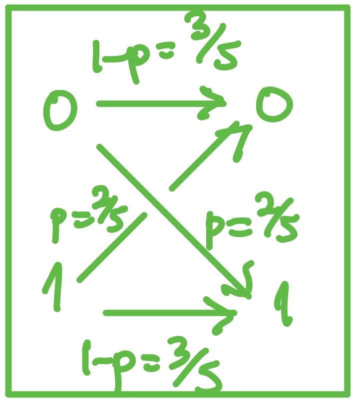
$$= \frac{P\left(\underset{\text{received}}{y} \,\middle|\, \underset{\text{sent}}{x_i}\right) \cdot P\left(\underset{\text{sent}}{x_i}\right)}{P\left(\underset{\text{received}}{y}\right)}$$

$$= \frac{p^{d(y,x_i)}(1-p)^{\ell-d(y,x_i)} \cdot p_i}{P\left(\underset{\text{received}}{y}\right)}$$

Pick x_i maximizing this numerator

EXAMPLE What if in previous example we had these source probabilities?:

	p_i	
"OK"	$1/2$	$x_1 = 0000$
various distress signals?	$1/6$	$x_2 = 0101$
	$1/6$	$x_3 = 1010$
	$1/6$	$x_4 = 1111$



receiving $y = 0111$

Then the ideal observer maximizes the numerator in

$$P\left(\begin{matrix} x_i \\ \text{sent}\end{matrix} \,\middle|\, \begin{matrix} 0111 \\ \text{received}\end{matrix}\right) = \frac{\left(\frac{2}{5}\right)^{d(x_i,0110)} \cdot \left(\frac{3}{5}\right)^{4-d(x_i,0110)} \cdot p_i}{P\left(\begin{matrix}0111\\ \text{received}\end{matrix}\right)}$$

This numerator is

$$\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^1 \cdot \frac{1}{2} = \boxed{\frac{12}{625} \text{ for } x_1 = 0000}$$

$$\left(\frac{2}{5}\right)^1\left(\frac{3}{5}\right)^3 \cdot \frac{1}{6} = \frac{9}{625} \text{ for } x_2 = 0101$$

$$\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^1 \cdot \frac{1}{6} = \frac{4}{625} \text{ for } x_3 = 1010$$

$$\left(\frac{2}{5}\right)^1\left(\frac{3}{5}\right)^3 \cdot \frac{1}{6} = \frac{9}{625} \text{ for } x_4 = 1111$$

The ideal observer decoding

Channel capacity & Shannon's Noisy Coding Theorem (§4.4, 4.5)

As with avglength(f) for u.d. codes, is there a limit on the rate $\frac{\log_2(|\mathcal{C}|)}{\ell}$ of a choice of length ℓ binary codewords $\mathcal{C} \subset \{0,1\}^*$ being sent through a noisy channel if we would like the probability of undetected error to be made arbitrarily small?

EXAMPLE

Note that if we began with codewords being all 2^ℓ words w_i of length ℓ in $\{0,1\}^*$, and encoding them with the r-fold repetition code

$\mathcal{C} = \{w_1 w_2 \cdots w_r\} \subset \{0,1\}^*$ with words of length $r\ell$

then the max error probability $\longrightarrow 0$ as $r \to \infty$

but also

$$\text{rate}(\mathcal{C}) = \frac{\log_2(|\mathcal{C}|)}{r\ell} = \frac{\log_2(2^\ell)}{r\ell} = \frac{1}{r} \longrightarrow 0 \text{ as } r \to \infty.$$

Q: Can we do better with the rate, still having error probability $\to 0$?

Yes, and how much better again relates to quantifying information/entropy of the source $X=\{x_1,\dots,x_m\}$ probs $p_1,\dots,p_m$

and the new source $Y=\{y_1,\dots,y_n\}$, that has probabilities calculable from the channel probabilities matrix $p_{ij} := P(y_j \text{ rec'd} \mid x_i \text{ sent})$

$$\begin{array}{cc} p_1 & x_1 \\ \vdots & \vdots \\ p_m & x_m \end{array} \rightsquigarrow \boxed{\begin{array}{c}\text{channel} \\ C \\ (p_{ij})\end{array}} \rightsquigarrow \begin{array}{c} y_1 \\ \vdots \\ y_n \end{array} \qquad P(y_j \text{ rec'd}) = \sum_{i=1}^{m} p_{ij}\, x_i$$

For each event $\{y_j \text{ received}\}$, one can calculate $P\left(\begin{matrix} x_i \\ \text{sent}\end{matrix} \middle| \begin{matrix} y_j \\ \text{received}\end{matrix}\right)$ for $j = 1, 2, \ldots, n$ and use it to define ...

DEF'N: The conditional entropies

$$H\left(X \middle| \begin{matrix} y_j \\ \text{received}\end{matrix}\right) := \sum_{i=1}^{m} P\left(\begin{matrix} x_i \\ \text{sent}\end{matrix} \middle| \begin{matrix} y_j \\ \text{rec'd}\end{matrix}\right) \log_2\left(\frac{1}{P\left(\begin{matrix} x_i \\ \text{sent}\end{matrix} \middle| \begin{matrix} y_j \\ \text{rec'd}\end{matrix}\right)}\right)$$

and then the entropy of X given Y,

$$H(X|Y) = \sum_{j=1}^{n} P\left(\begin{matrix} y_j \\ \text{rec'd}\end{matrix}\right) H\left(X \middle| \begin{matrix} y_j \\ \text{rec'd}\end{matrix}\right)$$

and finally the information about X given by Y

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

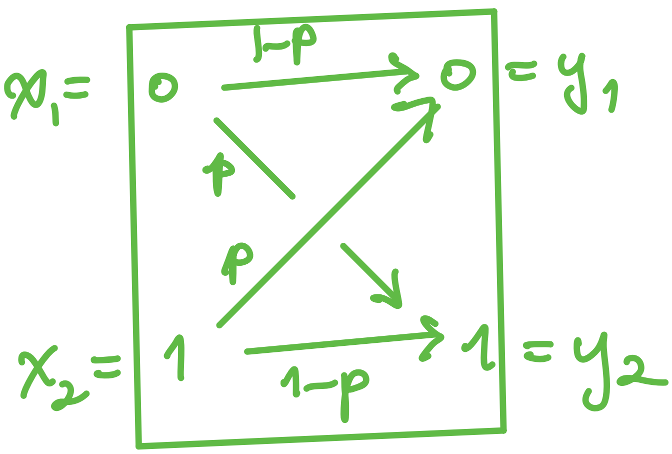
That is, we expect that despite the noise, knowing Y should decrease our surprise about X, by $I(X|Y)$ bits.

Finally, we can define...

DEF'N: The channel capacity of C

$$\text{capacity}(C) := \max \left\{ I(X|Y) : \begin{array}{l} \text{source probabilities} \\ p_1, \dots, p_m \text{ for} \\ X = \{x_1, \dots, x_m\} \end{array} \right\}$$

EXAMPLE Garrett calculates with some easy calculus that the BSC with error prob p



has

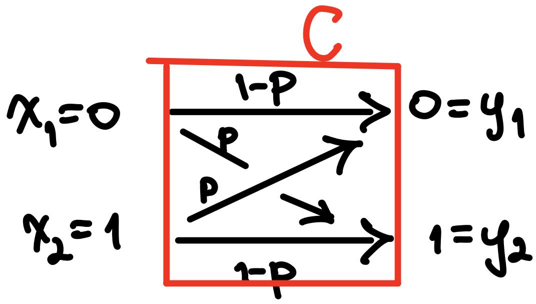
$$\text{capacity}(C) = 1 + p\log_2(p) + (1-p)\log_2(1-p)$$

(and max of $I(X|Y)$ is achieved for $P(x_1) = P(x_2) = \frac{1}{2}$, regardless of the BSC error probability p)

A peek inside that calculation of cap(C) for BSC...

$P(X=x_1)=r$

$P(X=x_2)=1-r$



leads to (with some straightforward algebra)

$$f(r) := I(X|Y) = H(X) - H(X|Y)$$
$$= p\log_2(p) + (1-p)\log_2(1-p) - \left[A\log_2 A + B\log_2 B\right]$$

where $A = r(1-p)+(1-r)p$
$B = rp + (1-r)(1-p)$

Want to compute

$$\text{cap}(C) = \max_{\substack{(p_1,\,p_2)\\ \overset{\shortparallel}{r}\quad \overset{\shortparallel}{1-r}}} I(X|Y) = \max_r f(r)$$

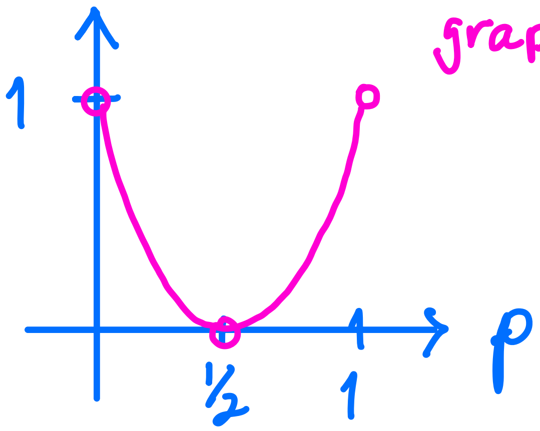
so set $0 = f'(r) = \frac{d}{dr}f(r)$, and find that it is maximized when $r = 1/2$, regardless of p.

But then when $r = 1/2$, $A = \frac{1}{2}(1-p) + (1-\frac{1}{2})p = \frac{1}{2}(1-p+p) = \frac{1}{2}$

$B = \frac{1}{2}p + (1-\frac{1}{2})(1-p) = \frac{1}{2}(p+1-p) = \frac{1}{2}$

and $\text{cap}(C) = p\log_2(p) + (1-p)\log_2(1-p) - \left[\frac{1}{2}\log_2(\frac{1}{2}) + \frac{1}{2}\log_2(\frac{1}{2})\right]$

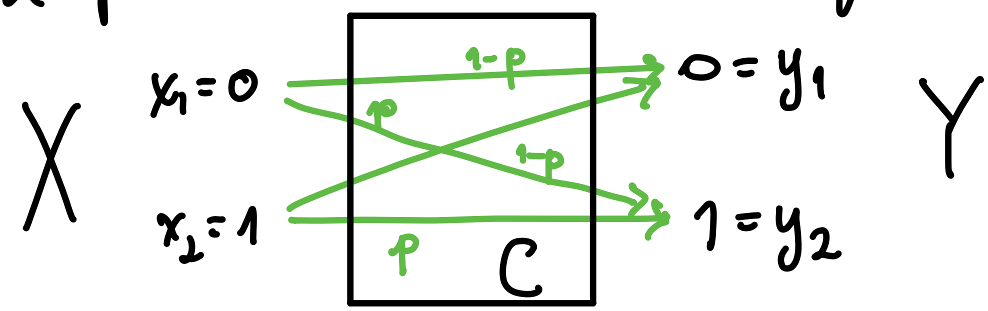
$= p\log_2(p) + (1-p)\log_2(1-p) + 1$



graph of $y = \text{capacity}(\text{BSC with error prob } p)$
$= 1 + p\log_2(p) + (1-p)\log_2(1-p)$

EXAMPLE Note BSC with error prob $p = 1/2$ has capacity $1 + \frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1 - \frac{1}{2} - \frac{1}{2} = 0$

It's a special case of this family of useless channels:



where one can check that for any choice of source probabilities $X = \{x_1, x_2\}$, one has
probs: p_1 p_2

X, Y independent $\Rightarrow H(X \mid y_j \text{ occurs}) = H(X) \ \forall y_j \in Y$

$\Rightarrow H(X|Y) = H(X)$

i.e. $I(X|Y) = 0$

$\Rightarrow \text{capacity}(C) = 0.$

No way to detect errors, even with long repetition codes and very low rates!

Shannon's Noisy Coding Thm (§4.5)

Let C be a memoryless channel, and pick any R in the range $0 < R < \text{capacity}(C)$.

Then one can find a sequence of codes

$$C_n \subset \{0,1\}^* \text{ for } n = 1, 2, 3, \dots$$

with

- C_n consists of words of length n
- $\text{rate}(C_n) \to R$ as $n \to \infty$
- using max likelihood (= min Hamming distance) decoding, the max probability of a word in C_n being decoded wrong $\to 0$ as $n \to \infty$.

Roman §3.4.4 states it, but both he and Garrett prove it only for the BSC with error probability p.

An interesting feature of the proof is, using fairly easy probabilistic estimates, one can pick $\mathcal{C}_n$ with high probability to be

$2^{\lfloor R \cdot n \rfloor}$ randomly chosen (!) words

of length n

$$\left(\text{so } \operatorname{rate}(\mathcal{C}_n) = \frac{\lfloor R \cdot n \rfloor}{n} \longrightarrow R \text{ as } n \to \infty\right)$$

DRAWBACK: Efficiently doing minimum-distance decoding with $\mathcal{C}$ chosen randomly is hard.

Roman also proves an accompanying result:

THEOREM (Weak Converse to Shannon's Noisy Coding Thm.)

[Roman Thm. 3.3.6]

Let C be a discrete memoryless channel, and pick any $R > \mathrm{cap}(C)$.

Let $\mathcal{C}_n \subset \{0,1\}^n$ for $n = 1, 2, 3, \ldots$ be any sequence of length n binary codes that have $|\mathcal{C}_n| \approx 2^{nR}$

(so that $\mathrm{rate}(\mathcal{C}_n) = \frac{\log_2(|\mathcal{C}_n|)}{n} \approx \frac{\log_2(2^{nR})}{n} = \frac{nR}{n} = R$).

Then assuming codewords from $\mathcal{C}_n$ are chosen uniformly at random, when passing them through C and doing minimum distance decoding,

$$\left(\begin{array}{c}\text{probability}\\ \text{of error}\\ \text{in decoding}\\ \text{from } \mathcal{C}_n\end{array}\right) \geq 1 - \frac{1}{nR} - \frac{\mathrm{cap}(C)}{R}$$

$$\Big\downarrow \text{ as } n \to \infty$$

$$1 - \frac{\mathrm{cap}(C)}{R} > 0.$$

BAD!

since $R > \mathrm{cap}(C)$