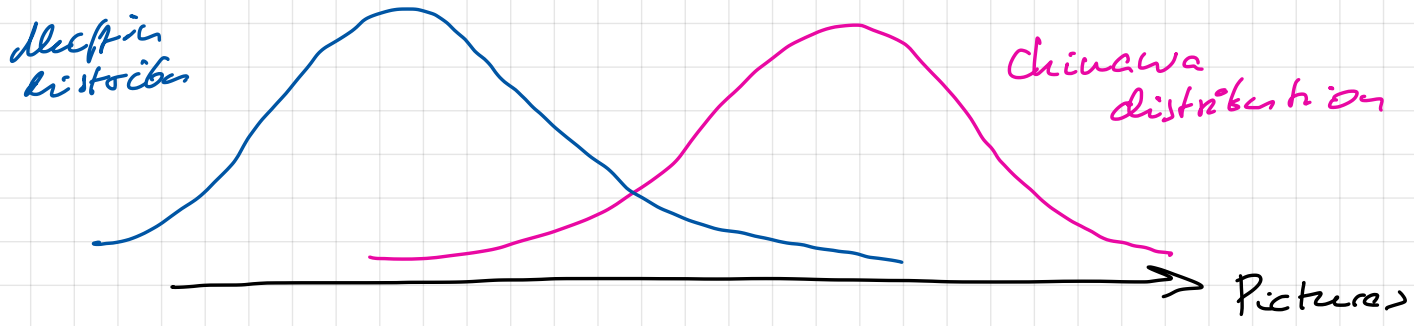


x is the picture
c is the category of the picture (chihuahua or muffin).

We have conditional probabilities:

$p(x \mid c)$ probability to see x given the category

$P(x \mid \text{chihuahua})$

$P(x \mid \text{muffin})$

And we want to know $P(c \mid x)$.

Bayes Formula links $P(c \mid x)$ to $P(x \mid c)$ using $P(c)$ and $P(x)$

$P(c)$: Probability of the category

$P(x)$: Probability of the picture.

To recap: CH channel

$CH(x)$ is random

$$P_{x \to y} = P(CH(x) = y).$$

If $X \in_R C \subseteq W$ (we have a random input X in the code C, independent of the noise of channel), then we have

$Y = CH(X)$, which is a random variable whose distribution depends not only on $P_{x \to y}$ but also on the distribution of X.

$$P(X=x \mid Y=y) \overset{\text{Bayes}}{=} \frac{P(Y=y \mid X=x)}{P(Y=y)} \cdot P(X=x)$$

$$= \frac{P_{x \to y}}{\sum_{x' \in C} P(Y=y \mid X=x') \cdot P(X=x')} \, P_x$$

$$= \frac{P_{x \to y}}{\sum_{x' \in C} P_{x' \to y} \cdot P_{x'}} \cdot P_x$$

Once you know p_x ($x \in C$) and $P_{x \to y}$ (for all x and y) you can calculate

$P(X=x \mid Y=y)$ and decide how you decode.

For a given y, you pick $x \in C$ with the highest conditional probability $P(X=x \mid Y=y)$.

A decoder $D: W \to C$ acting by this rule is called maximal likelihood decoder or shortly an ML decoder.

A common choice for p_x is $p_x = \frac{1}{|C|}$, which means that the random codeword is uniformly distributed on the code C. In this case,

$$P(X=x \mid Y=y) = \frac{p_{x \to y}}{\sum_{x' \in C} p_{x' \to y}}$$

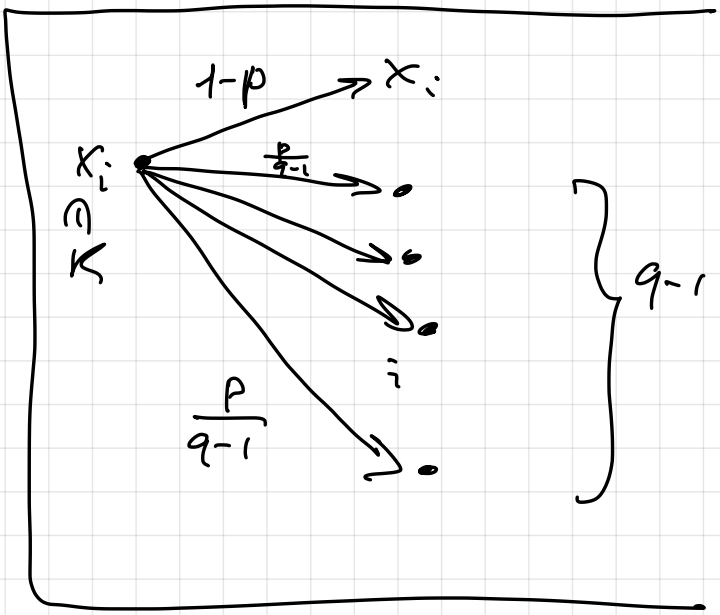
How do we decode based on the values $p_{x \to y}$ how to decode?

$\sum_{x' \in C} p_{x' \to y}$ does not depend on x. So, an ML decoder picks, for a given y, an $x \in C$ with the biggest value of $p_{x \to y}$.

On the other hand, for a code $C \subseteq K^n$ a decoder $D: K^n \to C$ is called a minimum distance decoder if it decodes $y \in K^n$ to a codeword $D(y) \in C$ closest to x in the Hamming distance.

Prop 3.35 Consider a q-ary symmetric channel on K^n with the cross-over probability $p \in (0, 1-\frac{1}{q})$ $(q \in \mathbb{N}, q \geq 2)$ and a random input for the channel independent on the noise of the channel and uniformly distributed in a given code $C \subseteq K^n$, $C \neq \emptyset$. In this setting, $D: K^n \to C$ is an ML decoder if and only if D is a minimum distance decoder.

Proof:



$$p_{x \to y} = (1-p)^{n-d(x,y)} \left(\frac{p}{q-1}\right)^{d(x,y)}$$

$$= (1-p)^n \cdot \left(\frac{p}{(q-1)(1-p)}\right)^{d(x,y)}$$

An ML decoder maximizes $p_{x \to y}$ over $x \in C$ for a given $y \in K^n$. But this corresponds to

$$\max_{x \in C} \left(\frac{p}{(q-1)(1-p)}\right)^{d(x,y)},$$ where

$$0 < \frac{p}{(q-1)(1-p)} < 1 \iff \frac{p}{1-p} < q-1$$

$$\iff \frac{1-p}{p} > \frac{1}{q-1}$$

$$\iff \frac{1}{p} > 1 + \frac{1}{q-1}$$

$$\iff p < \frac{1}{1+\frac{1}{q-1}}$$

$$p < \frac{q-1}{q-1+1}$$

$$p < \frac{q-1}{q}$$

$$p < 1 - \frac{1}{q}.$$

So, maximization of $P_{x \to y}$ for $x \in C$ is equivalent to minimization of $d(x,y)$ for $x \in C$. □

3.4. Shannon's theorem on the capacity of noisy channels.

Def. 3.36 The rate of a code $C \subseteq K^n$, $C \neq \emptyset$ over a q-ary alphabet K ($n, q \in \mathbb{N}$, $q \geq 2$) is defined by

$$R(C) = \frac{\log_q |C|}{n}$$

is called the information rate, or simply the rate of the code C.

Example 3.37 $C = \{00000, 01010, 10101, 11110\} \subseteq \{0,1\}^5$.

$|C| = 4$

$n = 5$

$q = 2$

$\log_q |C| = 2$

$$R(C) = \frac{2}{5} = 0.4$$

Example 3.38

$$C = \{\underbrace{(a, a, \ldots, a)}_{n \text{ times}} : a \in K\} \subseteq K^n$$

K is a q-ary alphabet.

$$R(C) = \frac{\log_q |C|}{n} = \frac{\log_q q}{n} = \frac{1}{n}$$

Remark 3.39 $R(C)$ actually only depends on the size of C and K^n.

$$R(C) = \frac{\log_q |C|}{\log_q |K^n|}$$ now you can write it as $$R(C) = \frac{\log_2 |C|}{\log_2 |K^n|}$$ or as

$$R(C) = \log_{|K^n|} |C| .$$

Def 3.40 Consider a random variable X distributed on a finite set S. We define a q-ary entropy of X (for $q \in \mathbb{N}$ $q \geq 2$) as

$$H_q(X) = -\sum_{x \in S} P(X=x) \log_q P(X=x).$$

where the terms with $P(X=x)=0$ are interpreted as zero.

Example 3.41 Fair coin $X \in_R \{0,1\}$

$q=2$

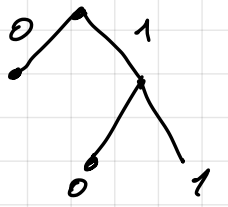
$$H_2(X) = -P(X=0)\log_2 P(X=0) - P(X=1)\log_2 P(X=1)$$

$$= -\frac{1}{2}\cdot(-1) - \frac{1}{2}\cdot(-1) = 1.$$

Two independent tosses of a fair coin $X \in_R \{0,1\}^2$.

$$H_2(X) = -\Big(P(X=00)\log_2 P(X=00)$$
$$+ P(X=01)\log_2 P(X=01)$$
$$+ P(X=10)\log_2 P(X=10)$$
$$+ P(X=11)\log_2 P(X=11)\Big)$$

$$= -4\cdot\frac{1}{4}\cdot\log_2\frac{1}{4} = 2.$$



If head, stop
If tale, another toss

$X \in_R \{0, 10, 11\}$

$$P(X=0) = \frac{1}{2}$$
$$P(X=10) = \frac{1}{4}$$
$$P(X=11) = \frac{1}{4}$$

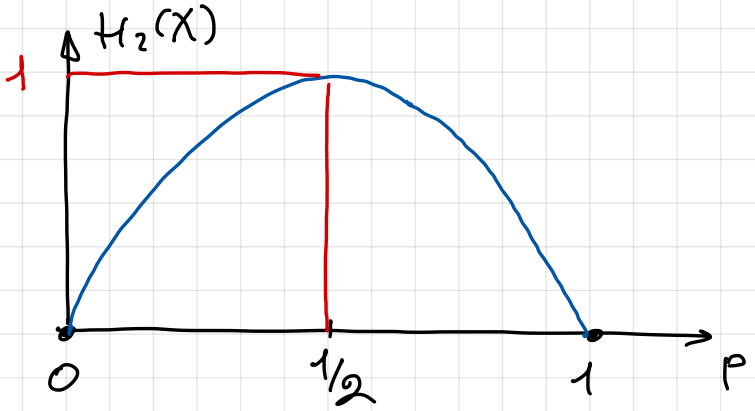
$$H_2(X) = -\left(\frac{1}{2}\cdot\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$= \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2$$

$$= 1.5.$$

Example 3.42 Binary entropy of a biased coin $X \in_R \{0,1\}$ with $p = P(X=1)$.

$$-H_2(X) = P(X=0)\log_2 P(X=0) + P(X=1)\log_2 P(X=1)$$
$$= (1-p)\log_2(1-p) + p\cdot\log_2 p.$$



Rem. 3.43

Now, let's look at the situation where we are interested in the entropy of a random vector $X=(X_1, X_2, \dots, X_n)$ with the independent components $X_1, \dots, X_n$ distributed in some finite sets $X_i \in_R S_i$.

$$P(X=(x_1,x_2,\dots,x_n)) = P(X_1=x_1 \text{ and } X_2=x_2 \dots \text{ and } X_n=x_n)$$
$$\overset{\text{by independence}}{=\!=} P(X_1=x_1)\,P(X_2=x_2)\cdot\dots\cdot P(X_n=x_n)$$

So, the q-ary entropy of X is

$$H_q(X) =$$

$$\sum_{\substack{x_1 \in S_1 \\ x_2 \in S_2 \\ \vdots \\ x_n \in S_n}} P(X_1=x_1)\,P(X_2=x_2)\cdot\dots\cdot P(X_n=x_n)\log_q \prod_{i=1}^{n} P(X_i=x_i)$$

$$= \sum_{x_1 \in S_1, \dots, x_n \in S_n} P(X_1=x_1)\,P(X_2=x_2)\dots P(X_n=x_n)\log_q P(X_i=x_i)$$
$$i=1,\dots,n$$

Reality check: $S_1 = \{0,1\}$, $S_2 = \{0,1\}$

$n = 2$

- $H_q(X) = P(X_1=0)P(X_2=0)\log_q P(X_1=0)P(X_2=0)$
$+ P(X_1=0)P(X_2=1)\log_q P(X_1=0)P(X_2=1)$
$+ P(X_1=1)P(X_2=0)\log_q P(X_1=1)P(X_2=0)$
$+ P(X_1=1)P(X_2=1)\log_q P(X_1=1)P(X_2=0).$

$\overset{?}{=} -(H_q(X_1) + H_q(X_2))$

$= P(X_1=0)\log_q P(X_1=0) + P(X_1=1)\log_q P(X_1=1)$
$+ P(X_2=0)\log_q P(X_2=0) + P(X_2=1)\log_q P(X_1=1)$

Let's see: $\log_q P(X_1=0)P(X_2=0)$
$= \log_q P(X_1=0) + \log_q P(X_2=0)$

$\rightsquigarrow P(X_1=0)P(X_2=0)\log_q P(X_1=0)P(X_2=0)$
$= P(X_1=0)P(X_2=0)\log_q P(X_1=0) +$
$P(X_1=0)P(X_2=0)\log_q P(X_2=0).$

$p_i(0) = P(X_i=0)$
$p_i(1) = P(X_i=1)$

$p_1(0)p_2(0)\log_q p_1(0) + p_1(0)p_2(0)\log_q p_2(0)$
$p_1(0)p_2(1)\log_q p_1(0) + p_1(0)p_2(1)\log_q p_2(1)$
$p_1(1)p_2(0)\log_q p_1(1) + p_1(1)p_2(0)\log_q p_2(0)$
$p_1(1)p_2(1)\log_q p_1(1) + p_1(1)p_2(1)\log_q p_2(1)$

$$= \underbrace{p_1(0)\log_q p_1(0) + p_1(1)\log_q p_1(1)}_{H_q(X_1)} + \underbrace{p_2(0)\log_q p_2(0) + p_2(1)\log_q p_2(1)}_{H_q(X_2)}$$

In general:

$$-H_q(X) =$$

$$\underbrace{\left(\sum_{x_2 \in S_2} P(X_2 = x_2)\right)}_{1}\underbrace{\left(\sum_{x_3 \in S_3} P(X_3 = x_3)\right)}_{1}\cdots\underbrace{\left(\sum_{x_n \in S_n} P(X_n = x_n)\right)}_{1}\underbrace{\left(\sum_{x_1 \in S_1} P(X_1 = x_1)\log_q P(X_1 = x_1)\right)}_{H_q(X_1)}$$

$$+ \vdots +$$

$$\underbrace{\left(\sum_{x_1 \in S_1} P(X_1 = x_1)\right)}_{1}\underbrace{\left(\sum_{x_2 \in S_2} P(X_2 = x_2)\right)}_{1}\cdots\underbrace{\left(\sum_{x_{n-1} \in S_{n-1}} P(X_{n-1} = x_{n-1})\right)}_{1}\underbrace{\left(\sum_{x_n \in S_n} P(X_n = x_n)\log_q P(X_n = x_n)\right)}_{H_q(X_n)}$$

Let's summarize. If $X_1, \ldots, X_n$ are independent. Then

$$H_q(X_1, \ldots, X_n) = H_q(X_1) + \ldots + H_q(X_n).$$

Rem 3.44. Let's calculate the entropy of the noise from the q-ary symmetric channel.

$CH(x_1,\dots,x_n) = (y_1,\dots,y_n)$

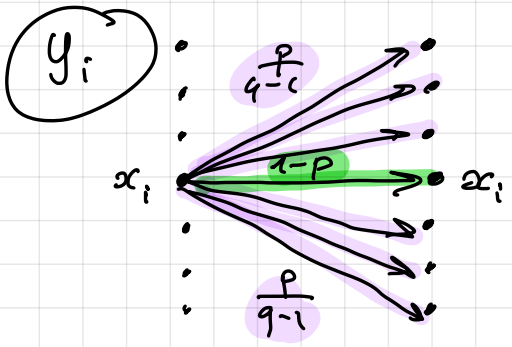
with independent $y_1,\dots,y_n \in_R K$

with $|K| = q$ $(q \in \mathbb{N},\ q \geq 2)$.

We want to know $H_q(y_1,\dots,y_n)$.

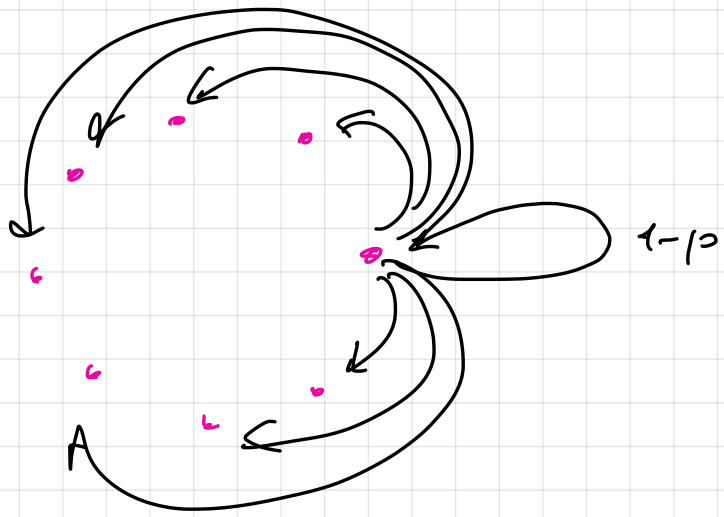
Since $y_1,\dots,y_n$ are independent, we have

$$H_q(y_1,\dots,y_n) = H_q(y_1) + \dots + H_q(y_n).$$
$$= n \cdot H_q(y_1).$$



$$-H_q(y_1) = (1-p)\log_q(1-p)$$
$$(q-1)\frac{p}{q-1}\log_p\frac{p}{q-1}$$

$$H_q(y_1,\dots,y_n) = n\left(-(1-p)\log_q(1-p)\right.$$
$$\left.-p\log_q\frac{p}{q-1}\right)$$



Now we need to compare two things against one another.

$$R(C) = \frac{\log_q |C|}{n}$$

rate of the code $C \subseteq K^n$ $(q = |K|)$.

against the entropy of the noise

$$H_q(Y_1, \dots, Y_n) = n \cdot \left(-(1-p)\log_q(1-p) - p \log_q \frac{1-p}{q-1}\right).$$

Effectively, $c \in C$ carries $\log_q |C|$ symbols of the alphabet.

On average, the noise kills about $H_q(Y_1, \dots, Y_n)$ symbols of the code.

We surely want that

$$n - \log_q |C| \geq H_q(Y_1, \dots, Y_n)$$

If we divide by n, we can write as $1 - R(C) \geq H_q(Y_1)$.

Shannon's theorem asserts that, if we take a slightly stronger inequality than the one above, there exists a code with the desired rate such that the decoding for this code works most with a high probability.

The rate $R(C) \leq 1 - H_q(y_1)$ is the limit and we get close to that limit.

Theorem 3.45 Consider the communication over a symmetric channel with the q-ary alphabet $q \in \mathbb{N}, q \geq 2$ and the cross-over probability $p \in (0, 1 - \frac{1}{q})$.

For every $\varepsilon \in (0,1)$ there exists a code length $n \in \mathbb{N}$ and a code $C \subseteq K^n$ of that length such that the rate of C is

$$R(C) \geq 1 + (1-p)\log_q(1-p) + p\log_q \frac{p}{q-1} - \varepsilon$$

and the minimum distance decoder $D: K^n \to C$ returns the correct codeword with the probability at least $1 - \varepsilon$.