Definition 0.1. *Let X be a random variable, values in set $\mathcal{X}$ (finite)*

$$p_X(x) = \mathbb{P}[X = x]$$

for $x \in \mathcal{X}$.

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) = \mathbb{E}(-\log_2 p(X))$$

where $p = (p_1, \ldots, p_n)$ is a probability vector ($p_i \geq 0, p_1 + \cdots + p_n = 1$). Then

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

This is the entropy of the random variable or its distribution (which are equivalent).

Uniform distribution:

$$H(U_N) = H(\underbrace{\frac{1}{N}, \ldots, \frac{1}{N}}_{n \text{ vectors}}) = \log_2 N$$

is the entropy of uniform distribution on N elements.

Proposition 0.2 (Properties)**.** *1. $H(X)$ depends only on the probabilities and not on any specific interpretation, i.e. if $\mathcal{X}'$ another set and $f : \mathcal{X} \to \mathcal{X}'$ bijective and $X' = f(X)$ then $H(X) = H(H'$ and $p_{X'}(f(x)) = p_X(x)$.*

2. For fixed n, $(p_1, \ldots, p_n) \mapsto H(p_1, \ldots, p_n)$ is continuous. In particular, $p_1 \mapsto H(p_1, 1 - p_1)$ is continuous on $[0, 1]$, maximum for $p_1 = 1 - p_1 = \frac{1}{2}$ is $H(\frac{1}{2}, \frac{1}{2}) = 1$, Back to original case: also here the maximum for $p_1 = \cdots = p_n = \frac{1}{n}$

add pic

If (X, Y) is a pair of random variables, X values in $\mathcal{X}$ and Y are values in $\mathcal{Y}$, both finite. Let $Z = (X, Y)$ with values in $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. We write the joint entropy $H(X, Y) = H(Z)$ as above.

$$H(X, Y) = -\sum_{x,y} p_{X,Y}(x, y) \log_2 p_{X,Y}(x, y)$$

0.1 Conditional entropy

Remark (Recall)**.** *(a) Marginal distribution of X and Y.*

$$p_X(x) = \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y)$$

$$p_Y(y) = \sum_{x \in \mathcal{X}} p_{X,Y}(x, y)$$

(b) Conditional distribution of Y, given $X = x$

$$p(y|x) = p_{(Y|X=x)}(y) = \mathbb{P}[Y = y | X = x] = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

where $p_X(x) > 0$ else if $p_X(x) = 0$ we have $p(x|y) = 0$. Note that

$$\sum_{y} p(y|x) = 1$$

The entropy of conditional distribution:

$$-\sum_{y\in\mathcal{Y}} p(y|x)\log_2 p(y|x) = H(Y|X=x)$$

Definition 0.3 (and Lemma: Conditional entropy)**.**

$$\begin{aligned}
H(Y|X) &\coloneqq \sum_{x\in\mathcal{X}} p_X(x)H(Y|X=x)\\
&= -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log_2 p(y|x)\\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)(\log_2 p(x,y) - \log_2 p_X(x))\\
&= H(X,Y) - (-\sum_x \underbrace{(\sum_y p(x,y))}_{=p_X(x)}\log_2 p_X(x)\\
&= H(X,Y) - H(X)
\end{aligned}$$

This equivalent to

$$H(X,Y) = H(X) + H(Y|X)$$

Example. *Let* $\Omega = \{1,\dots,N\} = \mathcal{X}$*. And the probability* $p(k) = \mathbb{P}(\{k\})$ *is given.*

$$\begin{aligned}
Z &= \sum_{k=1}^{N} k\mathbb{1}_{\{k\}}\\
Z(k) &= k\\
\mathbb{P}[Z=k] &= p_k\\
X &= \sum_{k=1}^{N-2} k\mathbb{1}_{\{k\}} + (N-1)\mathbb{1}_{\{N-1,N\}}\\
X(k) &= k,\ \mathit{for}\ k\le N-2\\
X(N-1) = X(N) &= N-1
\end{aligned}$$

Define $Y = \mathbb{1}_{\{N\}}$*. Then*

$$\begin{aligned}
Y(k) &= 0,\ \mathit{for}\ k\le N-1\\
Y(N) &= 1
\end{aligned}$$

Now compute

$$\begin{aligned}
\mathbb{P}[X=k] &= p_k\ \mathit{for}\ k\le N-2\\
\mathbb{P}[X=N-1] &= p_{N-1}+p_N\\
\mathbb{P}[Y=1] &= p_N\\
\mathbb{P}[Y=0] &= 1-p_N
\end{aligned}$$

Consider the pair (X,Y)*:*
values in $\{(1,0),\dots,(N-2,0),(N-1,0),(N-1,1)\} = \mathcal{Z}$

probabilities $p_1, \dots, p_{N-2}, p_{N-1}, p_N$

$$\begin{aligned} f : \mathcal{Z} &\to \mathcal{X} \\ (k,l) &\mapsto k + l \end{aligned}$$

is a bijection. Other approach: $Z = X + Y$, *then we get the same bijection. computation:*

$$\begin{aligned} H(Z) &= -\sum_{k=1}^{N} p_k \log_2 p_k = H(X,Y) \\ H(X) &= -\sum_{k=1}^{N-2} p_k \log_2 p_k - (p_{N-1} + p_N) \log_2(p_{N-1} + p_N) \\ H(Y|X) &= \sum_{k=1}^{N-2} p_k H(Y|X = k) + (p_{N-1} + p_N) H(Y|X = N-1) \end{aligned}$$

$$\begin{aligned} \mathbb{P}[Y = 0|X = N-1] &= \frac{\mathbb{P}[Y = 0, X = N-1]}{\mathbb{P}[X = N-1]} = \frac{p_{N-1}}{p_{N-1} + p_N} \\ \mathbb{P}[Y = 1|X = N-1] &= \frac{p_N}{p_{N-1} p_N} \end{aligned}$$

$H(Y|X = k) = 0$ *for* $k \le N-2$ *since in this case* $p(Y) = 0$ *and the entropy of constant values is* 0.
We want to conclude a formula:

$$H(Y|X) = (p_{N-1} + p_N) H(\frac{p_{N-1}}{p_{N-1} + p_N}, \frac{p_N}{p_{N-1} + p_N})$$

$$H(p_1, \dots, p_N) = H(p_1, \dots, p_{N-2}, p_{N-1} + p_N) + (p_{N-1} + p_N) H\left(\frac{p_{N-1}}{p_{N-1} + p_N}, \frac{p_N}{p_{N-1} + p_N}\right) \tag{1}$$

Theorem 0.4 (Axiomatic definition of entropy)**.** *Let* $p = \{(p_1, \dots, p_N) : N \in \mathbb{N}, p_k \ge 0, \sum_{k=1}^{N} p_k = 1\}$. *Assume that* $H : p \to \mathbb{R}$ *is a function with*

1. *H is permutation (transposition) invariant:*

$$H(p_1, \dots, p_i, \dots, p_j, \dots, p_N) = H(p_1, \dots, p_j, \dots, p_i, \dots, p_N)$$

2. $p_0 \mapsto H(p_0, 1 - p_0)$ *is continuous on* $[0, 1]$

3. $H(\frac{1}{2}, \frac{1}{2}) = 1$

4. *Equation (1) holds whenever* $p_{N-1} + p_N > 0$. *Then*

$$H(p_1, \dots, p_N) = -\sum_{k=1}^{N} p_k \log_2 p_k$$

is an entropy.

Proof. for Math students next time $\square$

Theorem 0.5 (Chain rule)**.** *Let* $X_1, \ldots, X_n$ *be discrete random variables with values in* $\mathcal{X}$*. Then*

$$H(X_1, \ldots, X_n) = \sum_{k=1}^{n} \underbrace{H(X_k | X_{k-1}, \ldots, X_1)}_{=H(X_1) \text{ for } k=1}$$

Proof. Exercise:
Induction on n.
$H(X_1, \ldots, X_n) = H((X_1, \ldots, X_{N-1}), X_n)$ □