Your Paper

You

December 5, 2025

Abstract

Your abstract.

1 Entropy and Mutual information

Let $(\mathcal{X}, \mathcal{A}, \mathbb{P}_\mathcal{A})$ be a probability space. Let $X : \mathcal{X} \to \mathbb{R}$ be a discrete random variable on the space with probability density function $f_X : \mathcal{X} \to \mathbb{R}_+$. We use the shorthand notation $p(x) = \mathbb{P}_\mathcal{A}[X = x]$. Let $(\mathcal{Y}, \mathcal{B}, \mathbb{P}_\mathcal{B})$ be a probability space. Let $Y : \mathcal{Y} \to \mathbb{R}$ be a discrete random variable on the space with probability density function $f_Y : \mathcal{Y} \to \mathbb{R}_+$. We use the shorthand notation $p(y) = \mathbb{P}_\mathcal{B}[X = x]$.

Definition 1. Let X be a discrete random variable with distribution $p(x)$.
We define entropy as $H_q(X) = \mathbb{E}\left(-\log_q p(X)\right)$. Since entropy was originally defined in the context of compression by Shannon in TODO, we usually use $q = 2$ and $H(X) = \mathbb{E}\left(-\log_2 p(X)\right)$.

Theorem 1. *Existence of Entropy: If $\mathcal{X}$ is finite, $H_q(X)$ exists.*
Example, when Entropy does not exist: Let $X =$,

Proof. □

Convention: $0 \log(0) = 0$.
Using definition of expected value: $H_p(X) = \mathbb{E}\left(-\log_q(p(X)) = \mathbb{E}\left(\frac{1}{\log_q(p(X))}\right)\right)$.
Properties of Entropy:

Theorem 2. $H(X) \geq 0$.

Proof. Note that $\log(\frac{1}{[0,1]}) = \log([1, \infty]) = [0, \infty]$ and $p(X)(\mathcal{X}) \in [0, 1]$.
Using the monotonicity of the expected value, we obtain

$$0 \leq \mathbb{E}\left(\log\left(\frac{1}{p(X)}\right)\right) = -\mathbb{E}\left(\log\left(p(X)\right)\right) = H(X)$$

□

Definition 2. Let X, Y be discrete random variables with marginal distributions $p(x), p(y)$ and with joint distribution $p(x, y)$. We define the following:

$$\begin{aligned}
\text{Conditional Entropy:} \quad & H(X \mid Y) = -\mathbb{E}\left(\log p(X \mid Y)\right) \\
\text{Joint Entropy:} \quad & H(X, Y) = -\mathbb{E}\left(\log p(X, Y)\right) \\
\text{Relative Entropy:} \quad & D(p(x) \| q(x)) = \mathbb{E}_p\left(\log \frac{p(X)}{q(X)}\right) \text{ with the conventions...} \\
\text{Mutual information:} \quad & I(X; Y) := D(p(x, y) \| p(x)p(y))
\end{aligned}$$

Remark 1. From now on, we will always define these two discrete random variables X, Y from Definition 2.

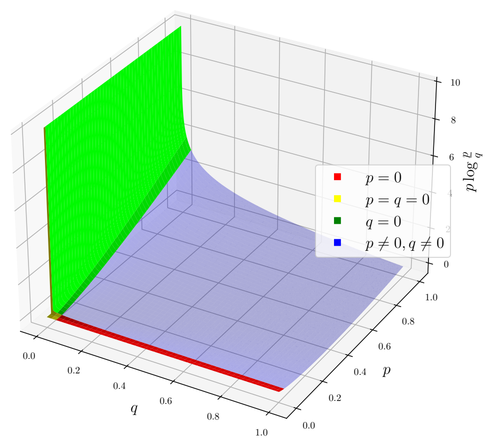
Remark 2. Under the assumptions of Definition 2, we have

$$\begin{aligned} D(p(x)\|q(x)) &= \mathbb{E}_p\left(\log\frac{p(X)}{q(X)}\right) && \text{(def. of relative entropy)} \\ &= \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} && \text{(def. of expected value)} \end{aligned}$$

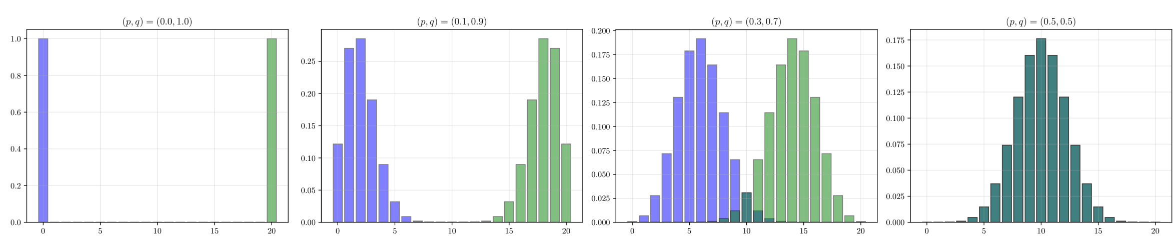
To understand the conventions, we can look at the limit cases:

1. **Case** $p \in (0,1], q = 0$**:** $\lim_{q\to 0^+} p\log\frac{p}{q} = \lim_{q\to 0^+}(p\log p - p\log q) = \infty$.
2. **Case** $p = 0, q \in (0,1]$**:** $0\log\frac{0}{q} = 0$.
3. **Case** $p = q = 0$**:** Case 1 logic yields $\lim_{q\to 0^+} p\log\frac{p}{q} = \infty$ and Case 2 logic yields $0\log\frac{0}{0} = 0$.

As we want $\sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}$ to sum over $x \in \mathcal{X}, p(x) > 0$, we choose the convention $0\log\frac{0}{0} = 0$. We can visualize the the pointwise relative entropy function $(p,q) \mapsto \log\frac{p}{q}$:



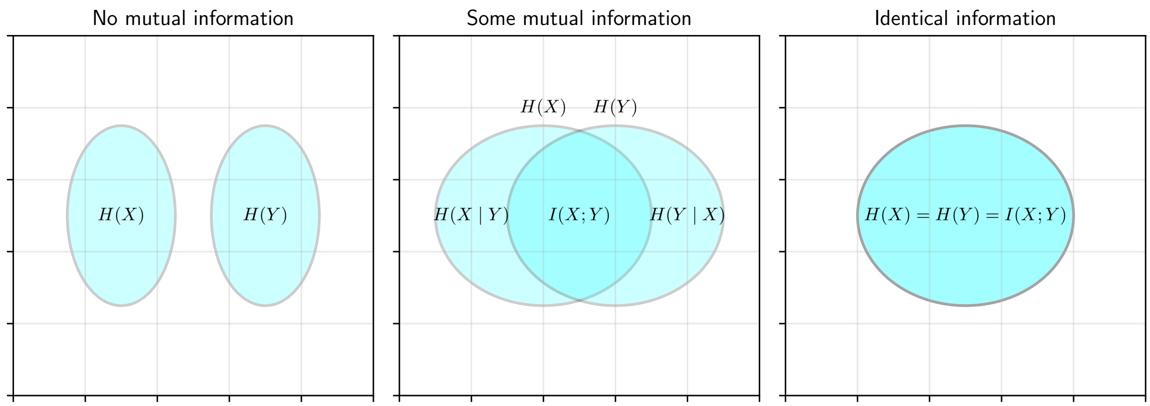
We can calculate the relative entropies for an example. The more alike the distributions are, the closer to zero the relative entropy is. Let $X \sim B(20,\alpha)$ and $Y \sim B(20,\beta)$ with $(\alpha,\beta) \in [0,1]^2$.



$$\begin{aligned} D(p(x)\|q(x)) &= \sum_{x=0}^{20} p(x)\log\frac{p(x)}{q(x)} \\ D(p(x)\|q(x)) &= 0\log\frac{0}{1} + 1\log\frac{1}{0} = 0 + \infty = \infty && (\alpha = 0, \beta = 1) \\ D(p(x)\|q(x)) &\approx 35.2 && (\alpha = 0.1, \beta = 0.9) \\ D(p(x)\|q(x)) &\approx 6.8 && (\alpha = 0.3, \beta = 0.7) \\ D(p(x)\|q(x)) &= \sum_{x=0}^{20} p(x)\log 1 = 0 && (\alpha = 0.5, \beta = 0.5) \end{aligned}$$

How do these concepts relate?
TODO: Add a venn diagram visualisation.

Remark 3. The relationship between entropy, conditional entropy and mutual information can be visualized:



Theorem 3. *We can formalise the visual insight from before:*

1. $I(X;Y) = H(Y) - H(Y|X)$
2. $I(X;Y) = I(Y;X)$
3. $I(Y;X) = H(X) - H(X|Y)$
4. $I(X;Y) = H(X,Y)$
5. $I(X;X) = H(X)$

Proof. 1. We can use the definition of mutual information and relative entropy to obtain:

$$
\begin{aligned}
I(X;Y) &= D(p(x,y) \| p(x)p(y)) && \text{(by def. of mutual info.)} \\
&= \mathbb{E}_{p(x,y)} \left(\log \frac{p(x,y)}{p(x)p(y)} \right) && \text{(by def. relative entropy)} \\
&= \mathbb{E}_{p(x,y)} \left(\log \frac{p(x)p(y|x)}{p(x)p(y)} \right) && \text{(using cond. probability)} \\
&= \mathbb{E}_{p(x,y)} \left(\log \frac{p(y|x)}{p(y)} \right) && \text{(simplify fraction)} \\
&= \mathbb{E}_{p(x,y)} (\log p(y|x)) - \mathbb{E}_p(x,y) (\log p(y)) && \text{(simplify logarithm)} \\
&= H(Y|X) - H(Y) && \text{(by def. of entropy)}
\end{aligned}
$$

2. The definition of mutual information yields:

$$
\begin{aligned}
I(X;Y) &= D(p(x,y) \| p(x)p(y)) && \text{(by def. of mutual info.)} \\
&= D(p(y,x) \| p(y)p(x)) && \text{(TODO: Idk)} \\
&= I(X;Y)
\end{aligned}
$$

3. Follows directly from 2 and 3.

4.

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(X|Y) && \text{(by 1)} \\
&= H(Y) - (H(X,Y) - H(X)) && \text{(chain rule)} \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

5. Using 1 we get $I(X;X) = H(X) - H(X|X) = H(X)$.

$\square$

Theorem 4. *The entropy function H(X) is concave.*

Proof. $\square$

2 Advanced properties and Inequalities

3 Axiomatic Definition