



## Why

### Definition

Consider two distributions on the same finite set. The *entropy* of the first distribution *relative* to the second distribution is the difference of the cross entropy of the first distribution relative to the second and the entropy of the second distribution. We call it the *relative entropy* of the first distribution with the second distribution. People also call the relative entropy the *Kullback-Leibler divergence* or *KL divergence*.

### Notation

Let  $A$  be a non-empty finite set. Let  $p : A \rightarrow \mathbf{R}$  and  $q : A \rightarrow \mathbf{R}$  be distributions. Let  $H(q, p)$  denote the cross entropy of  $p$  relative to  $q$  and let  $H(q)$  denote the entropy of  $q$ . The entropy of  $p$  relative to  $q$  is

$$H(q, p) - H(q).$$

Herein, we denote the entropy of  $p$  relative to  $q$  by  $d(q, p)$ .

### A similarity function

The relative entropy is a similarity function between distributions.

**Prop. 1.** *Let  $q$  and  $p$  be distributions on the same set. Then  $d(q, p) \geq 0$  with equality if and only if  $p = q$ .*

So,  $d$  has a few of the properties of a metric. However,  $d$  is not a metric; for example, it is not symmetric.

**Prop. 2.** *There exist distributions  $p : A \rightarrow \mathbf{R}$  and  $q : A \rightarrow \mathbf{R}$  (with  $A$  a non-empty finite set) such that*

$$d(q, p) \neq d(p, q).$$

### **Optimization Perspective**

If we want to find a distribution  $p : A \rightarrow \mathbf{R}$  to

$$\text{minimize } d(q, p),$$

then  $p = q$  is a solution.

