KL散度

KL散度来源于信息论。信息论主要研究如何量化数据中的信息。最重要的信息度量单位是熵Entropy，一般用H表示。分布的熵的公式如下：

只要稍加修改熵H的计算公式就能得到KL散度的计算公式。设p为观察得到的概率分布，q为另一分布来近似p，则p、q的KL散度为：

显然根据上面的公式，KL散度其实是数据的原始分布p和近似分布q之间的对数差值的期望。如果继续使用2为底的对数计算，则KL散度值表示信息损失的二进制位数。下面公式以期望表达KL散度：

一般，KL散度以下面的书写方式更常见：