交叉熵，相对熵：

熵的本质是香农信息量log(1/p)的期望。

那么一个样本的期望编码长度为。

错误分布q来表示来自真实分布p的平均编码长度

然后有：非真实分布q得到的平均编码长度H(p,q)大于根据真实分布p得到的平均编码长度H(p)。

q得到的平均编码长度比由p得到的平均编码长度多出的bit数称为“相对熵”

D(p||q)=H(p,q)-H(p)= 

它表示2个函数或概率分布的差异性：差异越大则相对熵越大，差异越小则相对熵越小，特别地，若2者相同则熵为0。注意，KL散度的非对称性。