# 背景

香农（Shannon）在A Mathematical Theory of Communication中提出来的。

# 什么是信息熵？

## 通俗理解

通俗地理解，一个事件的信息熵（Entropy）就是一个事件所包含的信息量。我们常常听到“这句话信息量好大”，比如“昨天花了10万，终于在西二环买了套四合院”。

这句话为什么信息量大？因为它的内容出乎意料，违反常理。由此引出：

* 可能性越低的事件，信息量越大。比如“我不会死”这句话信息量就很大。而确定事件的信息量就很小，比如“我是我妈生的”，信息量就很小甚至为0。
* 独立事件的信息量可以叠加。比如，事件“a. 张三今早吃了油条；b. 李四昨晚喝了豆浆”的信息量就恰好等于a+b的信息量。

## 更深入的例子

### 例1

假设甲从一个口袋里取球让乙来猜球的颜色，口袋里有一个蓝色的、一个红色的、一个绿色的, 一个橙色的。乙每次取球之后可以问一个问题，然后做出判断。

那么，如何让乙问最少的问题，得到正确答案呢？一个期望最好的策略如下：



图 1 例1的演绎

在不知道各种颜色的球的占比的情况下，只能认为各种球的占比是相同的。也就是说，甲拿到任何颜色的球的概率都是1/4。

由图1的演绎可知，乙提出的问题数恒为2。

### 例2

在例1的基础上，已知袋子中有1/8的球是绿的、1/8的球是橙的、1/4的球是红的、1/2的球是蓝的。这时，期望最优的策略如下：



图 2 例2的演绎

按照这个策略，甲拿蓝色球的时候，乙问1个问题；甲拿红色球的时候，乙问2个问题，甲拿绿色或者橙色球的时候，乙问3个问题。因此，猜中甲所拿的球，乙需要提出的问题数的期望为。

### 例3

现在，甲告诉乙：袋子里全都是蓝色的球。那么乙提出0个问题，就能猜中甲手里是什么球。

### 归纳

归纳上面的例子，可以发现，假设一种球出现的概率为，那么猜中该球所需要提问的次数是（数学证明略），事件“该球出现”的信息量为。一种通俗的理解方式就是，事件越稀有，信息量越大，而信息量越大的事件，就越需要更多的问题把它问出来。

因此，猜中一个随机拿的球，所需提问的次数的期望为



这就是**信息熵**的公式。

它的意义就是，在已知各种样本的出现概率的情况下，为一个随机样本进行分类所需的最优问题个数。