能量分布得越均匀，熵越大。

最大熵原理：不要把所有的鸡蛋放在一个篮子里，可以降低风险。

原理：保留全部的不确定性，将风险降到最小。

对于任何一组不自相矛盾的信息，这个最大熵模型不仅存在，而且是唯一的。

建议参考网址：<https://www.cnblogs.com/xfzhang/p/5916868.html>

1）知道骰子有几种（隐含状态数量），每种骰子是什么（转换概率），根据掷骰子掷出的结果（可见状态链），我想知道每次掷出来的都是哪种骰子（隐含状态链）。

这个问题呢，在语音识别领域呢，叫做解码问题。这个问题其实有两种解法，会给出两个不同的答案。每个答案都对，只不过这些答案的意义不一样。第一种解法求最大似然状态路径，说通俗点呢，就是我求一串骰子序列，这串骰子序列产生观测结果的概率最大。第二种解法呢，就不是求一组骰子序列了，而是求每次掷出的骰子分别是某种骰子的概率。比如说我看到结果后，我可以求得第一次掷骰子是D4的概率是0.5，D6的概率是0.3，D8的概率是0.2.第一种解法我会在下面说到，但是第二种解法我就不写在这里了，如果大家有兴趣，我们另开一个问题继续写吧。

2）还是知道骰子有几种（隐含状态数量），每种骰子是什么（转换概率），根据掷骰子掷出的结果（可见状态链），我想知道掷出这个结果的概率。

看似这个问题意义不大，因为你掷出来的结果很多时候都对应了一个比较大的概率。问这个问题的目的呢，其实是检测观察到的结果和已知的模型是否吻合。如果很多次结果都对应了比较小的概率，那么就说明我们已知的模型很有可能是错的，有人偷偷把我们的骰子給换了。

3）知道骰子有几种（隐含状态数量），不知道每种骰子是什么（转换概率），观测到很多次掷骰子的结果（可见状态链），我想反推出每种骰子是什么（转换概率）。

这个问题很重要，因为这是最常见的情况。很多时候我们只有可见结果，不知道HMM模型里的参数，我们需要从可见结果估计出这些参数，这是建模的一个必要步骤。