EM算法： 已知数据 找出模型的参数

背景：假设现在有两枚硬币1和2，随机抛掷后正面朝上概率分别为P1，P2。为了估计这两个概率，做实验，每次取一枚硬币，连掷5下，记录下结果，如下：



对于已知模型和数据，可以很容易算出参数

P1 = （3+1+2）/ 15 = 0.4  
P2= （2+3）/10 = 0.5

如果不知道属于哪个模型 只知道数据的话 就要用EM算法找参数



步骤1：将5个回合设为5个隐变量 z(1,2,3,4,5)

步骤2.1：先随便给P1和P2赋一个值，比如：

P1 = 0.2

P2 = 0.7

根据初始的P1和P2，可以计算出



步骤2.2：

第1轮中最有可能的是**硬币2**  
第2轮中最有可能的是**硬币1**  
第3轮中最有可能的是**硬币1**  
第4轮中最有可能的是**硬币2**  
第5轮中最有可能的是**硬币1**

以此为估计值，

现在我们知道模型和数据

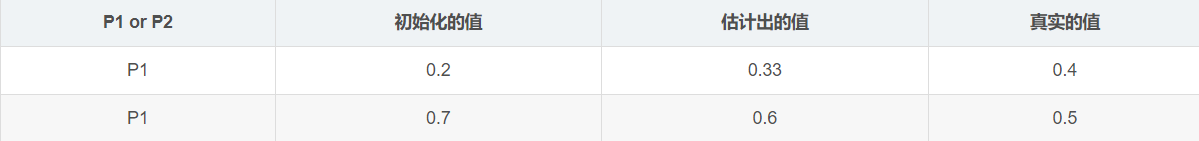
按MLE来估计下一轮的P1和P2

步骤3：

第一轮猜测为硬币2：3正-2反  
第一轮猜测为硬币1：2正-3反  
第一轮猜测为硬币1：1正-4反  
第一轮猜测为硬币2：3正-2反  
第一轮猜测为硬币1：2正-3反

新的P1,P2：

P1 = （2+1+2）/15 = 0.33  
P2=（3+3）/10 = 0.6



比较一轮循环后的P1P2和真实的P1P2

步骤4：循环2-3,只到更新幅度足够小