9.1

如例9.1的三硬币模型，假设观测数据不变，试选择不同的初值，例如，$\pi^{(0)}=0.46, p^{(0)}=0.55, q^{(0)}=0.67$，求模型参数为$\theta=(\pi,p,q)$的极大似然估计。

例9.1（三硬币模型） 假设有3枚硬币，分别记作A，B，C。这些硬币正面出现的概率分别是π，p和q。进行如下掷硬币试验：先掷硬币A，根据其结果选出硬币B或硬币C，正面选硬币B，反面选硬币C；然后掷选出的硬币，掷硬币的结果，出现正面记作1，出现方面记作0；独立地重复n次试验（这里，$n=10$），观测结果如下：
$1,1,0,1,0,0,1,0,1,1$
假设只能观测到掷硬币的结果，不能观测掷硬币的过程。

三硬币模型可以写作
$$\begin{aligned} P(y|\theta) &= \sum_z P(y,z|\theta) = \sum_z P(z|\theta)P(y|z,\theta) \\ &= \pi p^y (1-p)^{1-y} + (1-\pi) q^y (1-q)^{1-y} \end{aligned}$$

- 根据书中第9章的例9.1的三硬币模型的EM算法：

EM算法首先选取参数的初值，记作$\theta^{(0)}=(\pi^{(0)},p^{(0)},q^{(0)})$，然后通过下面的步骤迭代计算参数的估计值，直至收敛为止

第i次迭代参数的估计值为$\theta^{(i)}=(\pi^{(i)},p^{(i)},q^{(i)})$。EM算法的第$i+1$次迭代如下：

E步：计算在模型参数$\pi^{(i)}, p^{(i)}, q^{(i)}$下观测数据$y_j$来自掷硬币B的概率：

$$\mu_j^{(i+1)} = \frac{\pi^{(i)} (p^{(i)})^{y_j} (1-p^{(i)})^{1-y_j}}{\pi^{(i)} (p^{(i)})^{y_j} (1-p^{(i)})^{1-y_j} + (1-\pi^{(i)}) (q^{(i)})^{y_j} (1-q^{(i)})^{1-y_j}}$$

M步：计算模型参数的新估计值：

$$\pi^{(i+1)} = \frac{1}{n} \sum_{j=1}^N \mu_j^{(i+1)}$$

$$p^{(i+1)} = \frac{\displaystyle\sum_{j=1}^n \mu_j^{(i+1)} y_j}{\displaystyle\sum_{j=1}^n \mu_j^{(i+1)}}$$

$$q^{(i+1)} = \frac{\displaystyle\sum_{j=1}^n (1-\mu_j^{(i+1)}) y_j}{\displaystyle\sum_{j=1}^n (1-\mu_j^{(i+1)})}$$

采用编程方法直接计算解决

```
import math
```

```
class coin:
    def __init__(self,threshold=1e-6,max_iter=1000,theta):
    ##########################
    #    初始化模型参数        #
    #    threshold:收敛阈值    #
    #    max_iter:最大迭代次数  #
    #    theta:模型参数的初值   #
    ##########################
        self.threshold=threshold
        self.max_iter=max_iter
        self.pi,self.p,self.q=theta
    def mu(self,j):
    ##############################################
    #     （E步）计算mu                          #
    #    j:观测数据y的第j个                       #
    #    返回在模型参数下观测数据yj来自掷硬币B的概率 #
    ##############################################
        pro_1=self.pi*math.pow(self.p,data[j])*math.pow((1-self.p),1-data[j])
        pro_2=(1-self.pi)*math.pow(self.q,data[j])*math.pow((1-self.q),1-data[j])
        return pro_1/(pro_1+pro_2)
    def fit(self,data):
    ##############################################
    #    模型迭代                               #
    #    data:观测数据                           #
    ##############################################
        count=len(data)
        print("模型参数的初值：")
        print("pi={},p={},q={}".format(self.pi,self.p,self.q))
        print("EM算法训练过程：")
        for i in range(self.max_iter):
            #（E步）得到在模型参数下观测数据yj来自掷硬币B的概率
            _mu=[self.mu(j) for j in range(count)]
            #（M步）计算模型参数的新估计值
            pi=1/count*sum(_mu)
            p=sum([_mu[k]*data[k] for k in range(count)])/sum([_mu[k] for k in range(count)])
            q=sum([(1-_mu[k])*data[k] for k in range(count)])/sum([(1-_mu[k]) for k in range(count)])
            print('第{}次:pi={:.4f},p={:.4f},q={:.4f}'.format(i+1,pi,p,q))
            #计算误差值
            error=abs(self.pi-pi)+abs(self.p-p)+abs(self.q-q)
            self.pi=pi
            self.p=p
            self.q=q
            #判断是否收敛
            if error<self.threshold:
                print("模型参数的极大似然估计：")
                print("pi={:.4f},p={:.4f},q={:.4f}".format(self.pi,self.p,self.q))
                break
```

```
# 加载数据
data = [1, 1, 0, 1, 0, 0, 1, 0, 1, 1]
# 模型参数的初值
init_prob = [0.46, 0.55, 0.67]

# 三硬币模型的EM模型
em = ThreeCoinEM(prob=init_prob, tol=1e-5, max_iter=100)
# 模型训练
em.fit(data)
```

```
模型参数的初值:
pi=0.46,p=0.55,q=0.67
EM算法训练过程:
第1次:pi=0.4619,p=0.5346,q=0.6561
第2次:pi=0.4619,p=0.5346,q=0.6561
模型参数的极大似然估计:
pi=0.4619,p=0.5346,q=0.6561
```

9.2 证明定理9.2

定理 9.2 设 $L(\theta)=\log P(Y|\theta)$ 为观测数据的对数似然函数，$\theta^{(i)}(i=1,2,\cdots)$ 为 EM 算法得到的参数估计序列，$L(\theta^{(i)})(i=1,2,\cdots)$ 为对应的对数似然函数序列。

（1）如果 $P(Y|\theta)$ 有上界，则 $L(\theta^{(i)})=\log P(Y|\theta^{(i)})$ 收敛到某一值 L^*；

（2）在函数 $Q(\theta,\theta')$ 与 $L(\theta)$ 满足一定条件下，由 EM 算法得到的参数估计序列 $\theta^{(i)}$ 的收敛值 θ^* 是 $L(\theta)$ 的稳定点。

---

(1)：

- 由定理**9.1**，有$P(y|\theta^i)$是单调递增的，所以$L(\theta^i)$是单调递增的
- 由于$P(y|\theta^i)$是有上界的，所以$P(y|\theta^i)$有上界,所以$L(\theta^i)$有上界
- 由单调有界定理，有$L(\theta^i)$收敛

(2):

> 参考:Wu CF .J On the convergence properties of the EM algorithm. The Annals of Statistics,1983, 11:95-103

GLOBAL CONVERGENCE THEOREM. *Let the sequence* $\{x_k\}_{k=0}^{\infty}$ *be generated by* $x_{k+1}\in M(x_k)$, *where* M *is a point-to-set map on* X. *Let a solution set* $\Gamma\subset X$ *be given, and suppose that:* (i) *all points* x_k *are contained in a compact set* $S\subset X$; (ii) M *is closed over the complement of* Γ; (iii) *there is a continuous function* α *on* X *such that* (a) *if* $x\notin\Gamma$, $\alpha(y)>\alpha(x)$ *for all* $y\in M(x)$, *and* (b) *if* $x\in\Gamma$, $\alpha(y)\geq\alpha(x)$ *for all* $y\in M(x)$.

Then all the limit points of $\{x_k\}$ *are in the solution set* Γ *and* $\alpha(x_k)$ *converges monotonically to* $\alpha(x)$ *for some* $x\in\Gamma$.

> Gloabal Convergence Theorem: 全局收敛定理

THEOREM 1. *Let* $\{\phi_p\}$ *be a* GEM *sequence generated by* $\phi_{p+1} \in M(\phi_p)$, *and suppose that* (i) M *is a closed point-to-set map over the complement of* $\mathscr{S}$ (*resp.* $\mathscr{M}$), (ii) $L(\phi_{p+1}) > L(\phi_p)$ *for all* $\phi_p \notin \mathscr{S}$ (*resp.* $\mathscr{M}$).

Then all the limit points of $\{\phi_p\}$ *are stationary points* (*local maxima*) *of* L, *and* $L(\phi_p)$ *converges monotonically to* $L^* = L(\phi^*)$ *for some* $\phi^* \in \mathscr{S}$ (*resp.* $\mathscr{M}$).

- 对EM算法，$Q(\theta, \theta^i)$对θ，θ_i是连续的是closseness of M的充分条件,于是有$Theorem 2$

THEOREM 2. *Suppose* Q *satisfies the continuity condition* (10). *Then all the limit points of any instance* $\{\phi_p\}$ *of an* EM *algorithm are stationary points of* L *and* $L(\phi_p)$ *converges monotonically to* $L^* = L(\phi^*)$ *for some stationary point* ϕ^*.

- 对EM算法来说，容易说明$Q(\theta, \theta^i)$对θ，θ_i是连续,满足$Therorem 1(i)$的条件
- 由定理**9.1**，有$P(y|\theta^i)$是单调递增的,即$L(\theta^i)$是单调递增的,所以满足$Therorem 1(ii)$的条件
- 则自然满足Therorem 1的条件，同时得证在函数$Q(\theta, \theta')$与$L(\theta)$满足一定条件下，由EM算法得到的参数估计序列 θ^i的收敛值θ^*是$L(\theta)$的稳定点。

> PROOF. Since (10) is sufficient for (i) of Theorem 1, it remains to prove (ii) of Theorem 1 for all $\phi_p \notin \mathscr{S}$. Consider a ϕ_p, which is in the interior of Ω by (9). Since ϕ_p maximizes $H(\phi \mid \phi_p)$ over $\phi \in \Omega$ according to (4), $D^{10}H(\phi_p \mid \phi_p) = 0$. Therefore $DL(\phi_p) = D^{10}Q(\phi_p \mid \phi_p) \neq 0$ for any $\phi_p \notin \mathscr{S}$ from the definition of $\mathscr{S}$, implying that ϕ_p is not a local maximum of $Q(\phi \mid \phi_p)$ over $\phi \in \Omega$. From the definition of the M-step, $Q(\phi_{p+1} \mid \phi_p) > Q(\phi_p \mid \phi_p)$. Together with (4), this proves $L(\phi_{p+1}) > L(\phi_p)$ for all $\phi_p \notin \mathscr{S}$. The desired result follows. □