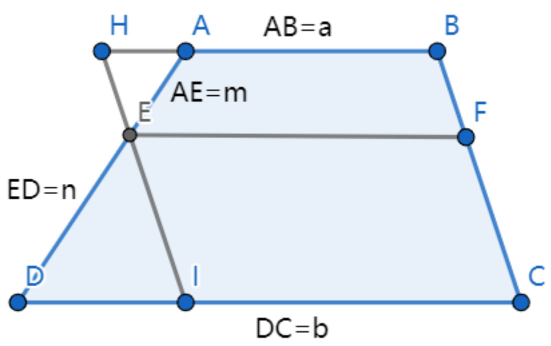
EM算法

1. 背景知识

1.1. 相似三角形

1. 定理：两角分别对应相等的两个三角形相似。
2. 定理：相似三角形任意对应线段的比等于相似比。

1.2. 梯形中位线推论



已知：在梯形ABCD中，$\mathrm{AB}//\mathrm{DC}$，$\mathrm{AB}=a, \mathrm{DC}=b$，E为边AB上的任意一点，$\mathrm{EF}//\mathrm{DC}$，且EF与BC相交与点F。

结论：$\mathrm{EF} = \dfrac{\mathrm{n}^*\mathrm{a} + \mathrm{m}^*\mathrm{b}}{\mathrm{m}+\mathrm{n}}$

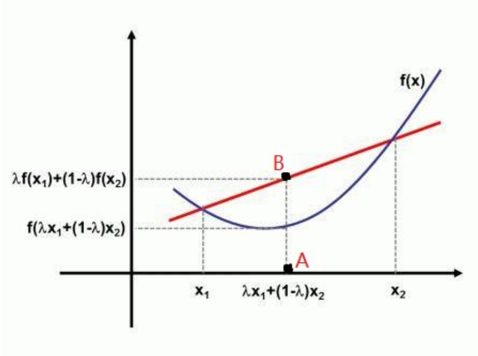
证明：过E做直线HI平行于BC，由相似三角形定理可知，$\triangle AEH \sim \triangle EDI$。

所以，$\dfrac{\mathrm{AE}}{\mathrm{ED}} = \dfrac{\mathrm{HA}}{\mathrm{DI}} = \dfrac{\mathrm{EF - a}}{\mathrm{b - EF}} = \dfrac{\mathrm{m}}{\mathrm{n}}$

化简可得：$\mathrm{EF} = \dfrac{\mathrm{n}^*\mathrm{a} + \mathrm{m}^*\mathrm{b}}{\mathrm{m}+\mathrm{n}}$

特别地，当$m+n=1$时，$\mathrm{EF} = \mathrm{n}^*\mathrm{a} + (1\text{-n})^*\mathrm{b}$

1.3. Jensen(琴生)不等式



若f是凸函数，则：

$$f\left(\lambda x_1+(1-\lambda) x_2\right) \leq \lambda f\left(x_1\right)+(1-\lambda) f\left(x_2\right) \tag{31}$$

其中，$\lambda \in [0,1]$。如果f是凹函数，则将$\leq$换为$\geq$即可。

证明1：令$x=\lambda x_1+(1-\lambda) x_2$，直线与曲线相交的两点分别为$(x_1, f(x_1))$，$(x_2, f(x_2))$。

根据直线两点式，可得直线方程为$\frac{f(x_2)-f(x_1)}{x_2-x_1}=\frac{f(x)-f(x_1)}{x-x_1}$

当$x=\lambda x_1+(1-\lambda) x_2$时，$y=\lambda f\left(x_1\right)+(1-\lambda) f\left(x_2\right)$

证明2：令$x=\lambda x_1+(1-\lambda) x_2$

因为$\lambda \in[0,1]$，所以$x \in[x_1, x_2]$。

因为$\frac{x-x_1}{x_2-x}=\frac{\lambda x_1+(1-\lambda) x_2-x_1}{x_2-\lambda x_1-(1-\lambda) x_2}=\frac{(1-\lambda)(x_2-x_1)}{\lambda(x_2-x_1)}=\frac{1-\lambda}{\lambda}$

根据梯形推导公式可得：

$y=\mathrm{AB}=\lambda f\left(x_1\right)+(1-\lambda) f\left(x_2\right)$

将上式中的λ推广到n个同样成立，即：

$$f\left(\lambda_1 x_1+\lambda_2 x_2+\ldots+\lambda_n x_n\right) \leq \lambda_1 f\left(x_1\right)+\lambda_2 f\left(x_2\right)+\ldots+\lambda_n f\left(x_n\right) \tag{32}$$

其中，$\lambda_1, \lambda_2, \ldots, \lambda_n \in[0,1],$

如果将t看做概率分布，则在概率论中：

$$f(\mathrm{E}[X]) \leq \mathrm{E}[f(x)] \tag{33}$$

其中，f是凸函数，X是随机变量，$E[X]$为X的期望

1.4. 期望

对随机变量x_i，对应的概率为p_i，则随机变量x的期望为：

$$E(x)=\sum_{i=1}^n x_i p_i \tag{34}$$

2. 三硬币模型引入

问题：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，出现反面记作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。问如何估计三硬币正面出现的概率，即三硬币模型的参数。

解：对每一次试验可如下建模

$$\begin{aligned}
P(y \mid \theta) & =\sum_z P(y, z \mid \theta) \\
& =\sum_z P(z \mid \theta) P(y \mid z, \theta) \\
& =P(z=1 \mid \theta) P(y \mid z=1, \theta)+P(z=0 \mid \theta) P(y \mid z=0, \theta) \\
& = \begin{cases}\pi p+(1-\pi) q, & \text { if } y=1 ; \\
\pi(1-p)+(1-\pi)(1-q), & \text { if } y=0 ;\end{cases} \\
& =\pi p^y(1-p)^{1-y}+(1-\pi) q^y(1-q)^{1-y}
\end{aligned} \tag{35}$$

其中，随机变量y是观测变量，表示一

则观测数据的似然函数为：

$$\begin{aligned}
P(Y \mid \theta) & =\sum_Z P(Y, Z \mid \theta) \\
& =\sum_Z P(Z \mid \theta) P(Y \mid Z, \theta) \\
& =\prod_{j=1}^n P\left(y_j \mid \theta\right) \\
& n
\end{aligned} \tag{36}$$

$$
= \prod_{j=1}^{n} \left[\pi p^{y_j} (1-p)^{1-y_j} + (1-\pi) q^{y_j} (1-q)^{1-y_j} \right]
$$

考虑求模型参数$\theta = (\pi, p, q)$的极大似然估计，即使用对数似然函数来进行参数估计，可得：

$$
\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} \ln P(Y \mid \theta) \\
&= \arg\max_{\theta} \ln \prod_{j=1}^{n} \left[\pi p^{y_j} (1-p)^{1-y_j} + (1-\pi) q^{y_j} (1-q)^{1-y_j} \right] \\
&= \arg\max_{\theta} \sum_{j=1}^{n} \ln \left[\pi p^{y_j} (1-p)^{1-y_j} + (1-\pi) q^{y_j} (1-q)^{1-y_j} \right]
\end{aligned} \tag{37}
$$

上式没有解析解，也就是没有办法直接通过求导方式解出(π, p, q)的值，只能使用迭代法进行求解。

3. EM算法

3.1. 为什么需要EM算法

概率模型有时候既含有观测变量又含有隐变量。如果概率模型的变量都是观测变量，那么给定数据，可以直接使用极大似然法估计或贝叶斯估计进行求解。但是，当模型含有隐变量时，就不能简单地使用这些估计方法。

EM算法就是解决含有隐变量的概率模型参数的极大似然估计。

3.2. EM算法推导

当面对含有隐变量的模型时，目标是极大化观测数据Y关于参数θ的对数似然函数，即极大化：

$$
L(\theta) = \ln P(Y \mid \theta) = \ln \sum_{Z} P(Y, Z \mid \theta) = \ln \left(\sum_{Z} P(Y \mid Z, \theta) P(Z \mid \theta) \right) \tag{38}
$$

注意到这一极大化的主要困难是上式中有未观测数据Z并有包含和（Z为离散型时）或 者积分（Z为连续型时）的对数。EM算法采用的是通过迭代逐步近似极大化 $L(\theta)$ 。

假设在第i次迭代后 θ 的估计值是 $\theta^{(i)}$, 我们希望新的估计值 θ 能使 $L(\theta)$ 增加，即 $L(\theta) > L\left(\theta^{(i)}\right)$ 并逐步达到极大值。为此，我们考虑两者的差：

$$
\begin{aligned}
L(\theta) - L\left(\theta^{(i)}\right) &= \ln \left(\sum_{Z} P(Y \mid Z, \theta) P(Z \mid \theta) \right) - \ln P\left(Y \mid \theta^{(i)}\right) \\
&= \ln \left(\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right)} \right) - \ln P\left(Y \mid \theta^{(i)}\right) \\
&\geq \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right)} - \ln P\left(Y \mid \theta^{(i)}\right) \\
&= \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right)} - 1 \cdot \ln P\left(Y \mid \theta^{(i)}\right) \\
&= \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right)} - \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \cdot \ln P\left(Y \mid \theta^{(i)}\right) \\
&= \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \left(\ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right)} - \ln P\left(Y \mid \theta^{(i)}\right) \right) \\
&= \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right) P\left(Y \mid \theta^{(i)}\right)}
\end{aligned} \tag{39}
$$

将$L\left(\theta^{(i)}\right)$移项，令：

$$
B\left(\theta, \theta^{(i)}\right) = L\left(\theta^{(i)}\right) + \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right) P\left(Y \mid \theta^{(i)}\right)} \tag{40}
$$

则：

$$
L(\theta) \geq B\left(\theta, \theta^{(i)}\right) \tag{41}
$$

即函数$B\left(\theta, \theta^{(i)}\right)$是$L(\theta)$的一个下界函数。

此时，若设θ^{i+1}使得$B\left(\theta, \theta^{(i)}\right)$达到极大，也就意味着：

$$
B\left(\theta^{(i+1)}, \theta^{(i)}\right) \geq B\left(\theta^{(i)}, \theta^{(i)}\right) \tag{42}
$$

由于：

$$
\begin{aligned}
B\left(\theta^{(i)}, \theta^{(i)}\right) &= L\left(\theta^{(i)}\right) + \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P\left(Y \mid Z, \theta^{(i)}\right) P\left(Z \mid \theta^{(i)}\right)}{P\left(Z \mid Y, \theta^{(i)}\right) P\left(Y \mid \theta^{(i)}\right)} \\
&= L\left(\theta^{(i)}\right) + \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P\left(Y, Z \mid \theta^{(i)}\right)}{P\left(Z, Y \mid \theta^{(i)}\right)} \\
&= L\left(\theta^{(i)}\right)
\end{aligned} \tag{43}
$$

进一步可推得：

$$
L\left(\theta^{(i+1)}\right) \geq B\left(\theta^{(i+1)}, \theta^{(i)}\right) \geq B\left(\theta^{(i)}, \theta^{(i)}\right) = L\left(\theta^{(i)}\right) \tag{44}
$$

即：

$$L\left(\theta^{(i+1)}\right) \geq L\left(\theta^{(i)}\right) \tag{45}$$

因此，任何能使得$B\left(\theta, \theta^{(i)}\right)$增大的$\theta$，也可以使$L(\theta)$增大。于是，问题就转化为求解能使得$B\left(\theta, \theta^{(i)}\right)$到达极大的$\theta^{i+1}$，即：

$$\begin{aligned}
\theta^{(i+1)} &= \arg \max_{\theta} B\left(\theta, \theta^{(i)}\right) \\
&= \arg \max_{\theta}\left(L\left(\theta^{(i)}\right)+\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right) P\left(Y \mid \theta^{(i)}\right)}\right) \\
&= \arg \max_{\theta}\left(\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln (P(Y \mid Z, \theta) P(Z \mid \theta))\right) \\
&= \arg \max_{\theta}\left(\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln P(Y, Z \mid \theta)\right) \\
&= \arg \max_{\theta} Q\left(\theta, \theta^{(i)}\right)
\end{aligned} \tag{46}$$

至此，完成了EM算法的一次迭代，求出的$\theta^{(i+1)}$作为下一次迭代的初始$\theta^{(i)}$。

综上所述，可以总结出EM算法的“E步”和“M步”分别为：

1. E步：导出Q函数

   计算完全数据的对数似然函数$\ln P(Y, Z \mid \theta)$关于给定观测数据Y和当前参数θ下对未观测数据Z的条件概率分布 $P\left(Z \mid Y, \theta^{(i)}\right)$的期望，也即Q函数：

$$Q\left(\theta, \theta^{(i)}\right)=E_{Z}\left[\ln P(Y, Z \mid \theta) \mid Y, \theta^{(i)}\right]=\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln P(Y, Z \mid \theta) \tag{47}$$

2. M步：Q函数极大

   求使得Q函数达到极大的$\theta^{(i+1)}$。

4. 使用EM求解三硬币模型

求解思路：

1. E步：导出Q函数
2. M步：求使得Q函数达到极大的 $\theta^{(i+1)}=\left(\pi^{(i+1)}, p^{(i+1)}, q^{(i+1)}\right)$

4.1. E步：导出Q函数

$$\begin{aligned}
Q\left(\theta \mid \theta^{(i)}\right) &= \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \ln P(Y, Z \mid \theta) \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \ln \left[\prod_{j=1}^{N} P\left(y_j, z_j \mid \theta\right)\right]\right\} \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right)\left[\sum_{j=1}^{N} \ln P\left(y_j, z_j \mid \theta\right)\right]\right\} \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right)\left[\ln P\left(y_1, z_1 \mid \theta\right)+\sum_{j=2}^{N} \ln P\left(y_j, z_j \mid \theta\right)\right]\right\} \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1 \mid \theta\right)+\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right)\left[\sum_{j=2}^{N} \ln P\left(y_j, z_j \mid \theta\right)\right]\right\} \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1 \mid \theta\right)\right\}+\sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right)\left[\sum_{j=2}^{N} \ln P\left(y_j, z_j \mid \theta\right)\right]\right\} \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1 \mid \theta\right)\right\}+ \\
&\quad \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_2, z_2 \mid \theta\right)\right\}+ \\
&\quad \cdots \quad + \\
&\quad \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_N, z_N \mid \theta\right)\right\}
\end{aligned} \tag{48}$$

单独考察 $\sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1 \mid \theta\right)\right\}$

$$\begin{aligned}
&\sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1 \mid \theta\right)\right\} \\
&= \sum_{z_1, z_2, \ldots, z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot P\left(z_1 \mid y_1, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1 \mid \theta\right)\right\} \\
&= \sum_{z_2, \ldots z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j, \theta^{(i)}\right) \cdot P\left(z_1=1 \mid y_1, \theta^{(i)}\right) \cdot \ln P\left(y_1, z_1=1 \mid \theta\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{z_2,\ldots,z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot P\left(z_1=0 \mid y_1,\theta^{(i)}\right)\cdot \ln P\left(y_1,z_1=0 \mid \theta\right)\right\}\\
&=P\left(z_1=1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1=1 \mid \theta\right)\sum_{z_2,\ldots,z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\right\}\\
&\quad+P\left(z_1=0 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1=0 \mid \theta\right)\sum_{z_2,\ldots z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\right\}\\
&=\left[P\left(z_1=1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1=1 \mid \theta\right)+P\left(z_1=0 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1=0 \mid \theta\right)\right]\sum_{z_2,\ldots,z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\right\}\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\sum_{z_2,\ldots,z_N}\left\{\prod_{j=2}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\right\}\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\left\{\sum_{z_3,\ldots,z_N}\left[\prod_{j=3}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot P\left(z_2=1 \mid y_2,\theta^{(i)}\right)\right]+\sum_{z_3,\ldots,z_N}\left[\prod_{j=3}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot P\left(z_2=0 \mid y\right.\right.\right.\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\left\{P\left(z_2=1 \mid y_2,\theta^{(i)}\right)\sum_{z_3,\ldots,z_N}\left[\prod_{j=3}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\right]+P\left(z_2=0 \mid y_2,\theta^{(i)}\right)\sum_{z_3,\ldots,z_N}\left[\prod_{j=3}^{N} P\left(z_j \mid y_j,\right.\right.\right.\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\left\{\left[P\left(z_2=1 \mid y_2,\theta^{(i)}\right)+P\left(z_2=0 \mid y_2,\theta^{(i)}\right)\right]\sum_{z_1,\ldots,z_N}\left[\prod_{j=3}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\right]\right\}\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\left\{\sum_{z_2} P\left(z_2 \mid y_2,\theta^{(i)}\right)\sum_{z_1,\ldots,z_N}\left[\prod_{j=3}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right]\right\}\right.\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\left\{\sum_{z_2} P\left(z_2 \mid y_2,\theta^{(i)}\right)\times\sum_{z_3} P\left(z_3 \mid y_3,\theta^{(i)}\right)\times\ldots\times\sum_{z_N} P\left(z_N \mid y_N,\theta^{(i)}\right)\right\}\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)\times\{1\times 1\times\ldots\times 1\}\\
&=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)
\end{aligned}
$$

所以，

$$
\sum_{z_1,z_2,\ldots,z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot\ln P\left(y_1,z_1 \mid \theta\right)\right\}=\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right) \tag{50}
$$

将其带回Q函数可得：

$$
\begin{aligned}
Q\left(\theta \mid \theta^{(i)}\right)=&\sum_{z_1,z_2,\ldots,z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot\ln P\left(y_1,z_1 \mid \theta\right)\right\}+\\
&\sum_{z_1,z_2,\ldots,z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot\ln P\left(y_2,z_2 \mid \theta\right)\right\}+\\
&\cdots\quad+\\
&\sum_{z_1,z_2,\ldots,z_N}\left\{\prod_{j=1}^{N} P\left(z_j \mid y_j,\theta^{(i)}\right)\cdot\ln P\left(y_N,z_N \mid \theta\right)\right\}\\
=&\sum_{z_1} P\left(z_1 \mid y_1,\theta^{(i)}\right)\ln P\left(y_1,z_1 \mid \theta\right)+\ldots+\sum_{z_N} P\left(z_N \mid y_N,\theta^{(i)}\right)\ln P\left(y_N,z_N \mid \theta\right)\\
=&\sum_{j=1}^{N}\left[\sum_{z_j} P\left(z_j \mid y_j,\theta^{(i)}\right)\ln P\left(y_j,z_j \mid \theta\right)\right]\\
=&\sum_{j=1}^{N}\left[P\left(z_j=1 \mid y_j,\theta^{(i)}\right)\ln P\left(y_j,z_j=1 \mid \theta\right)+P\left(z_j=0 \mid y_j,\theta^{(i)}\right)\ln P\left(y_j,z_j=0 \mid \theta\right)\right]
\end{aligned} \tag{51}
$$

由于：

$$
\begin{aligned}
&\begin{cases}P\left(y_j,z_j=1 \mid \theta\right)=\pi p^{y_j}(1-p)^{1-y_j}\\ P\left(y_j,z_j=0 \mid \theta\right)=(1-\pi)q^{y_j}(1-q)^{1-y_j}\end{cases}\\
&\begin{cases}P\left(z_j=1 \mid y_j,\theta^{(i)}\right)=\frac{P\left(z_j=1,y_j \mid \theta^{(i)}\right)}{P\left(y_j \mid \theta^{(i)}\right)}=\frac{\pi^{(i)}\left[p^{(i)}\right]^{y_j}\left(1-p^{(i)}\right)^{1-y_j}}{\pi^{(i)}\left[p^{(i)}\right]^{y_j}\left(1-p^{(i)}\right)^{1-y_j}+\left(1-\pi^{(i)}\right)\left[q^{(i)}\right]^{y_j}\left(1-q^{(i)}\right)^{1-y_j}}=\mu_j^{(i+1)}\\ P\left(z_j=0 \mid y_j,\theta^{(i)}\right)=1-P\left(z_j=1 \mid y_j,\theta^{(i)}\right)=\left(1-\mu_j^{(i+1)}\right)\end{cases}
\end{aligned} \tag{52}
$$

所以，Q函数的最终形式为：

$$
Q\left(\theta \mid \theta^{(i)}\right)=\sum_{j=1}^{N}\left\{\mu_j^{(i+1)}\ln\left[\pi p^{y_j}(1-p)^{1-y_j}\right]+\left(1-\mu_j^{(i+1)}\right)\ln\left[(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]\right\} \tag{53}
$$

4.2. 求Q函数达到极大的参数

该步骤求使得Q函数达到极大的$\theta^{(i+1)}=\left(\pi^{(i+1)},p^{(i+1)},q^{(i+1)}\right)$。

4.2.1. 对π求偏导

对Q函数关于π求一阶偏导数，并令一阶偏导数为0：

$$\begin{aligned}
\frac{\partial Q\left(\theta \mid \theta^{(i)}\right)}{\partial \pi} &= \sum_{j=1}^{N} \frac{\partial}{\partial \pi}\left\{\mu_j^{(i+1)} \ln\left[\pi p^{y_j}(1-p)^{1-y_j}\right] + \left(1-\mu_j^{(i+1)}\right) \ln\left[(1-\pi) q^{y_j}(1-q)^{1-y_j}\right]\right\} \\
&= \sum_{j=1}^{N}\left\{\mu_j^{(i+1)} \frac{p^{y_j}(1-p)^{1-y_j}}{\pi p^{y_j}(1-p)^{1-y_j}} + \left(1-\mu_j^{(i+1)}\right) \frac{-q^{y_j}(1-q)^{1-y_j}}{(1-\pi) q^{y_j}(1-q)^{1-y_j}}\right\} \\
&= \sum_{j=1}^{N}\left\{\frac{\mu_j^{(i+1)}(1-\pi) p^{y_j}(1-p)^{1-y_j} q_j(1-q)^{1-y_j}}{\pi(1-\pi) p^{y_j}(1-p)^{1-y_j} q^{y_j}(1-q)^{1-y_j}} + \frac{\left(\mu_j^{(i+1)}-1\right) \pi p_j(1-p)^{1-y_j} q^{y_j}(1-q)^{1-y_j}}{\pi(1-\pi) p^{y_j}(1-p)^{1-y_j} q^{y_j}(1-q)^{1-y_j}}\right\} \\
&= \sum_{j=1}^{N}\left\{\frac{\mu_j^{(i+1)} p^{y_j}(1-p)^{1-y_j} q^{y_j}(1-q)^{1-y_j} - \pi p^{y_j}(1-p)^{1-y_j} q^{y_j}(1-q)^{1-y_j}}{\pi(1-\pi) p^{y_j}(1-p)^{1-y_j} q^{y_j}(1-q)^{1-y_j}}\right\} \\
&= \sum_{j=1}^{N}\left[\frac{\mu_j^{(i+1)}-\pi}{\pi(1-\pi)}\right] \\
&= \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} - \sum_{j=1}^{N} \pi}{\pi(1-\pi)} \\
&= \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} - N\pi}{\pi(1-\pi)}
\end{aligned} \tag{54}$$

令上式为0，可得：

$$\begin{aligned}
\frac{\partial Q\left(\theta \mid \theta^{(i)}\right)}{\partial \pi} = \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} - N\pi}{\pi(1-\pi)} = 0 \\
\sum_{j=1}^{N} \mu_j^{(i+1)} - N\pi = 0 \\
N\pi = \sum_{j=1}^{N} \mu_j^{(i+1)} \\
\pi = \frac{1}{N} \sum_{j=1}^{N} \mu_j^{(i+1)} \Rightarrow \pi^{(i+1)} = \frac{1}{N} \sum_{j=1}^{N} \mu_j^{(i+1)}
\end{aligned} \tag{55}$$

4.2.2. 对$\mathbf{p}$求偏导

对Q函数关于$\mathbf{p}$求一阶偏导数，并令一阶偏导数为0：

$$\begin{aligned}
\frac{\partial Q\left(\theta \mid \theta^{(i)}\right)}{\partial p} &= \sum_{j=1}^{N} \frac{\partial}{\partial p}\left\{\mu_j^{(i+1)} \ln\left[\pi p^{y_j}(1-p)^{1-y_j}\right] + \left(1-\mu_j^{(i+1)}\right) \ln\left[(1-\pi) q^{y_j}(1-q)^{1-y_j}\right]\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial p}\left\{\mu_j^{(i+1)} \ln\left[\pi p^{y_j}(1-p)^{1-y_j}\right]\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial p}\left\{\mu_j^{(i+1)}\left[\ln \pi + y_j \ln p + (1-y_j) \ln(1-p)\right]\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial p}\left\{\mu_j^{(i+1)} \ln \pi + \mu_j^{(i+1)} y_j \ln p + \mu_j^{(i+1)}(1-y_j) \ln(1-p)\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial p}\left\{\mu_j^{(i+1)} y_j \ln p + \mu_j^{(i+1)}(1-y_j) \ln(1-p)\right\} \\
&= \sum_{j=1}^{N}\left\{\frac{\mu_j^{(i+1)} y_j}{p} + \frac{(-1) \cdot \mu_j^{(i+1)}(1-y_j)}{(1-p)}\right\} \\
&= \sum_{j=1}^{N} \frac{\mu_j^{(i+1)} y_j}{p} - \sum_{j=1}^{N} \frac{\mu_j^{(i+1)}(1-y_j)}{(1-p)} \\
&= \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} y_j}{p} - \frac{\sum_{j=1}^{N} \mu_j^{(i+1)}(1-y_j)}{(1-p)}
\end{aligned} \tag{56}$$

令上式等于0可得：

$$\begin{aligned}
\frac{\partial Q\left(\theta \mid \theta^{(i)}\right)}{\partial p} = \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} y_j}{p} - \frac{\sum_{j=1}^{N} \mu_j^{(i+1)}(1-y_j)}{(1-p)} = 0 \\
\frac{\sum_{j=1}^{N} \mu_j^{(i+1)} y_j}{p} = \frac{\sum_{j=1}^{N} \mu_j^{(i+1)}(1-y_j)}{(1-p)} \\
(1-p) \sum_{j=1}^{N} \mu_j^{(i+1)} y_j = p \sum_{j=1}^{N} \mu_j^{(i+1)}(1-y_j) \\
\sum_{j=1}^{N} \mu_j^{(i+1)} y_j - p \sum_{j=1}^{N} \mu_j^{(i+1)} y_j = p \sum_{j=1}^{N} \mu_j^{(i+1)} - p \sum_{j=1}^{N} \mu_j^{(i+1)} y_j \\
\sum_{j=1}^{N} \mu_j^{(i+1)} y_j = p \sum_{j=1}^{N} \mu_j^{(i+1)} \\
p = \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} y_j}{\quad} \Rightarrow p^{(i+1)} = \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} y_j}{\quad}
\end{aligned} \tag{57}$$

$$\sum_{j=1}^{N}\mu_j^{(i+1)} \qquad \sum_{j=1}^{N}\mu_j^{(i+1)}$$

4.2.3. 对q求偏导

对Q函数关于**q**求一阶偏导数，并令一阶偏导数为0：

$$
\begin{aligned}
\frac{\partial Q\left(\theta \mid \theta^{(i)}\right)}{\partial q} &= \sum_{j=1}^{N} \frac{\partial}{\partial q}\left\{\mu_j^{(i+1)} \ln \left[\pi p^{y_j}(1-p)^{1-y_j}\right]+\left(1-\mu_j^{(i+1)}\right) \ln \left[(1-\pi) q^{y_j}(1-q)^{1-y_j}\right]\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial q}\left\{\left(1-\mu_j^{(i+1)}\right) \ln \left[(1-\pi) q^{y_j}(1-q)^{1-y_j}\right]\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial q}\left\{\left(1-\mu_j^{(i+1)}\right)\left[\ln (1-\pi)+y_j \ln q+\left(1-y_j\right) \ln (1-q)\right]\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial q}\left\{\left(1-\mu_j^{(i+1)}\right) \ln (1-\pi)+\left(1-\mu_j^{(i+1)}\right) y_j \ln q+\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right) \ln (1-q)\right\} \\
&= \sum_{j=1}^{N} \frac{\partial}{\partial q}\left\{\left(1-\mu_j^{(i+1)}\right) y_j \ln q+\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right) \ln (1-q)\right\} \\
&= \sum_{j=1}^{N}\left\{\frac{\left(1-\mu_j^{(i+1)}\right) y_j}{q}+\frac{(-1) \cdot\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right)}{(1-q)}\right\} \\
&= \sum_{j=1}^{N} \frac{\left(1-\mu_j^{(i+1)}\right) y_j}{q}-\sum_{j=1}^{N} \frac{\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right)}{(1-q)} \\
&= \frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j}{q}-\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right)}{(1-q)}
\end{aligned}
\tag{58}
$$

令上式等于0可得：

$$
\begin{aligned}
\frac{\partial Q\left(\theta \mid \theta^{(i)}\right)}{\partial q}=\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j}{q}-\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right)}{(1-q)}=0 \\
\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j}{q}=\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right)}{(1-q)} \\
(1-q) \sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j=q \sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)\left(1-y_j\right) \\
\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j-q \sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j=q \sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)-q \sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j \\
\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j=q \sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) \\
q=\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j}{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)} \Rightarrow q^{(i+1)}=\frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j}{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)}
\end{aligned}
\tag{59}
$$

4.2.4. 总结

综上所述，为了求Q函数的极大值，可使用如下公式进行参数迭代：

$$
\begin{aligned}
\pi^{(i+1)} &= \frac{1}{N} \sum_{j=1}^{N} \mu_j^{(i+1)} \\
p^{(i+1)} &= \frac{\sum_{j=1}^{N} \mu_j^{(i+1)} y_j}{\sum_{j=1}^{N} \mu_j^{(i+1)}} \\
q^{(i+1)} &= \frac{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right) y_j}{\sum_{j=1}^{N}\left(1-\mu_j^{(i+1)}\right)}
\end{aligned}
\tag{60}
$$

5. 参考文档

1. 从最大似然到EM算法：一致的理解方式 - 科学空间|Scientific Spaces
2. (47条消息) EM算法之三硬币模型求解*zsdust*的博客-*CSDN*博客三硬币模型
3. 人人都懂EM算法 - 知乎 (zhihu.com)
4. EM算法详解 - 知乎 (zhihu.com)
5. 【机器学习】EM——期望最大（非常详细） - 知乎 (zhihu.com)