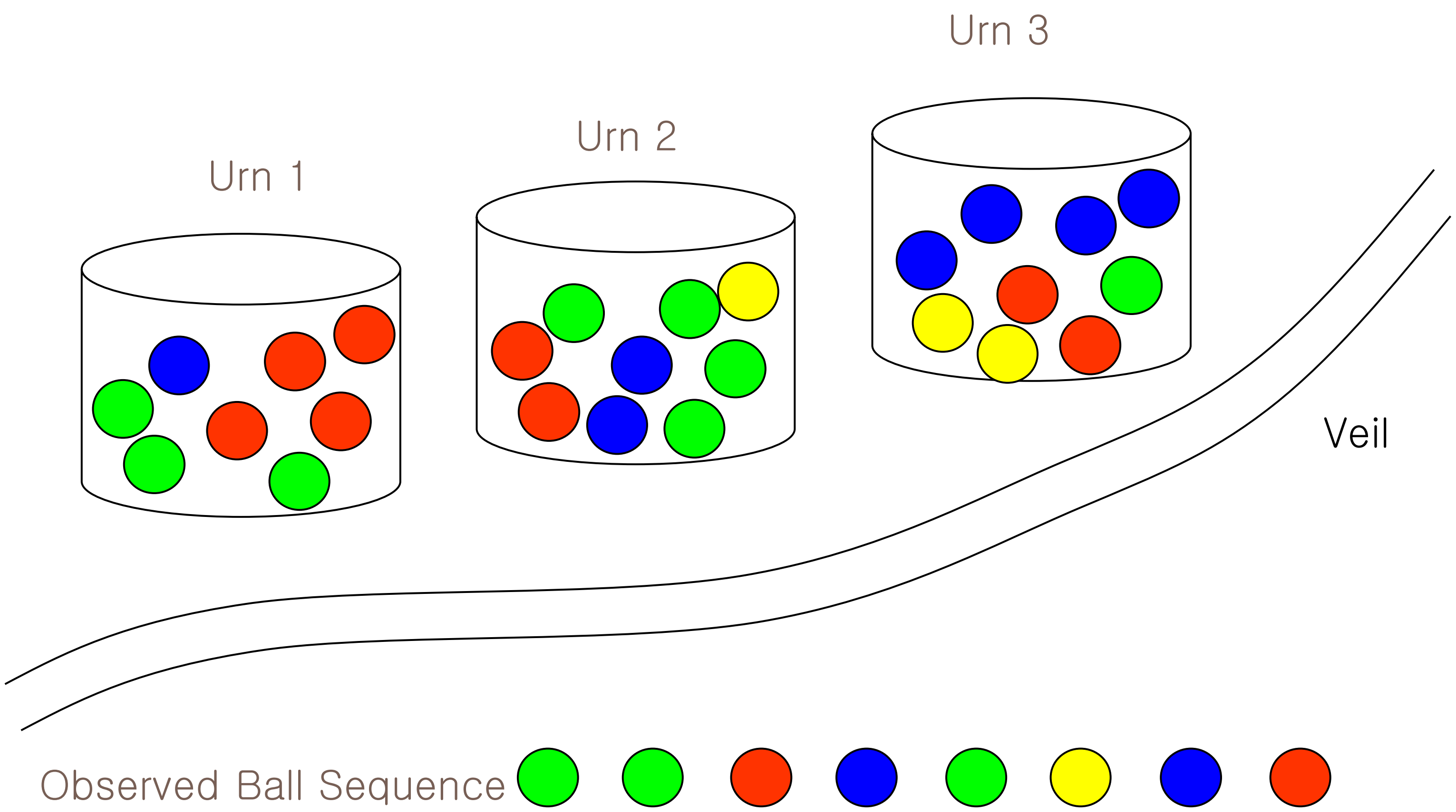
# 08-03 HMM 隐马尔可夫模型

# 真核基因组基因识别

- 真核基因识别较之原核基因识别要困难的多
  - 内含子
  - 调控因子
- HMM

# HMM实例



| 盒子 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 红球R | 5 | 3 | 6 | 8 |
| 白球W | 5 | 7 | 4 | 2 |

- 假设有4个盒子，每个盒子都有红白两种颜色的球。
- 按照如下规则取球：
- 从四个盒子中等概率取一个盒子；从跟这个盒子里随机抽取一个球。记录颜色并放回。
- 从当前盒子，随机转移到另外一个盒子。转移矩阵如下，

| | 盒1 | 盒2 | 盒3 | 盒4 |
|---|---|---|---|---|
| 盒1 | 0 | 1 | 0 | 0 |
| 盒2 | 0.4 | 0 | 0.6 | 0 |
| 盒3 | 0 | 0.4 | 0 | 0.6 |
| 盒4 | 0 | 0 | 0.5 | 0.5 |

- 确定转移的盒子之后，从中随机抽取1个球，记录颜色并放回。
- 如此重复5次。得到观测序列$O = \{R, W, R, W, R\}$

- 在这个过程中观察者只能观测到球的颜色，而观测不到盒子的序列。

- 在这个过程中，有两个随机序列：
  - 盒子的序列（状态序列）；隐藏的；
  - 球颜色的序列（观测序列）；可观测的。

- 这就是一个HMM的例子。

# HMM实例——约束

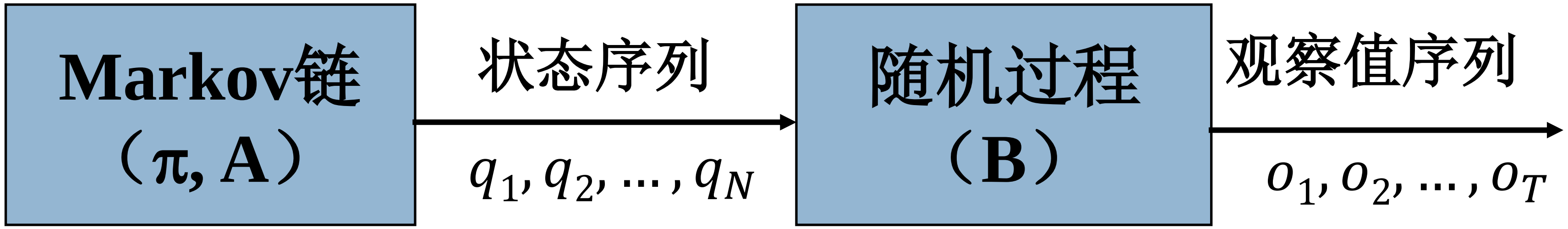
在上述实验中，有几个要点需要注意：

- **不能被直接观察盒子之间的转移**
- **从盒子中所选取的球的颜色和盒子并不是一一对应的**
- **每次选取哪个盒子由一组转移概率决定**

# HMM概念

- HMM的**状态（盒子）**是不确定或不可见的，只有通过观测序列（球的颜色）的随机过程才能表现出来
- 观察到的事件与状态并不是一一对应，而是通过一组概率分布相联系
- HMM是一个双重随机过程，两个组成部分：
  - **马尔可夫链**：描述状态的转移，用**转移概率**描述。
  - **一般随机过程**：描述状态与观察序列间的关系，用**观察值概率**描述。

# HMM组成




**HMM的组成示意图**

**盒1，盒2，**
**盒3，盒4，**

**R，R,W,W,R**

- 观测序列：O ={红，白，红}
- 状态集合（盒子对应状态）：$Q = \{$盒1，盒2，盒3，盒4$\}, N = 4$
- 观测集合（球的颜色）：$V = \{$红, 白$\}$, $M = 2$
- 状态序列和观测序列长度$T = 5$.
- 初始概率分布$\pi = (0.25, 0.25, 0.25, 0.25)^T$
- 状态转移矩阵：$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}_{N \times N}$
- 观测概率分布 $B = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}_{N \times M}$

# HMM的基本要素

设观测序列$O = (o_1, o_2, .., o_T)$,对应的状态序列$Q = (q_1, q_2, \dots, q_T)$

| 参数 | 含义 | 实例 | |
|---|---|---|---|
| N | 状态数目 | 盒子的数目 | $S = (S_1, S_2, \dots, S_N)$ |
| M | 每个状态可能的观察值数目 | 彩球颜色数目 | $V = (v_1, v_2, \dots, v_M)$ |
| A | 与时间无关的状态转移概率矩阵 | 在选定某个盒子的情况下，选择另一个盒的概率 | $A = [a_{ij}]_{N\times N}$<br>$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$ |
| B | 给定状态下，观察值概率分布 | 每个盒中的颜色分布 | $B = [b_j(k)]_{N\times M}$<br>$b_j(k) = P(o_t = v_k \mid q_t = S_j)$,<br>$1 \le j \le N, 1 \le k \le M$ |
| π | 初始状态空间的概率分布 | 初始时选择某盒的概率 | $\pi = (\pi_1, \dots, \pi_N)$ |

- 模型可用五元组$\lambda$=（$N, M, \pi$，$A$，$B$）或三元组 $\lambda = (\pi，A，B)$来描述HMM。

# HMM的两个假设

**设观测序列$O = (o_1, o_2, .., o_T)$,对应的状态序列$Q = (q_1, q_2, \ldots, q_T)$**

- 假设隐藏的Markov链在任意时刻$t$状态，只依赖前一时刻的状态，和其他时刻的状态和观测序列无关，和$t$也无关。（齐次Markov性）
  - $P(q_t|q_{t-1}, o_{t-1}, \ldots, q_1, o_1) = P(q_t|q_{t-1})$
- 假设任意时刻的观测只依赖于该时刻的Markov链的状态，与其他观测和状态无关。(观测独立性)
  - $P(o_t|q_T, o_T, q_{T-1}, o_{T-1}, \ldots, q_{t+1}, o_{t+1}, q_{t-1}, o_{t-1}, \ldots, q_1, o_1) = P(o_t|q_t)$

- **算法1：生成观测序列**
- 输入：HMM模型参数$\lambda = (A, B, \pi)$,观测序列长度$T$
- 输出：观察序列$O = (o_1, o_2, .., o_T)$

1. 按照初始状态分布$\pi$产生状态$S_i$
2. 令$t = 1$
3. 按照状态$q_t$的观测概率分布$b_{q_t}(k)$生成$o_t$
4. 按照状态$q_t$的状态转移概率$a_{q_t,q_{t+1}}$产生状态$q_{t+1}$,
5. 令$t = t + 1$, 如果$t < T$,转3.否则终止。

# HMM的三个基本问题

- 问题1：概率计算问题 （评估模型与观测序列之间匹配度）
  - 给定模型$\lambda=(A,B,\pi)$，观察序列$O=(o_1,o_2,..,o_T)$，计算在模型$\lambda$下观测序列$O$，出现的概率 $P(O|\lambda)$?
- 问题2：学习问题： （如何训练模型能更好地描述观测数据）
  - 已知观测序列$O=(o_1,o_2,..,o_T)$，估计模型$\lambda=(A,B,\pi)$参数，使得在该参数下，观测序列概率$P(O|\lambda)$最大。即，用极大似然估计方法估计参数。
- 问题3：预测问题或解码问题： （根据观测序列推断隐藏的状态序列）
  - 已知模型$\lambda=(A,B,\pi)$和观测序列$O=(o_1,o_2,..,o_T)$，求最有可能对应的状态序列$Q=(q_1,q_2,\ldots,q_T)$。即，在该状态序列下，观测序列的条件概率$P(O|Q)$最大?

# 问题1：概率计算问题：

给定模型 $\lambda = (A, B, \pi)$ ，观察序列 $O = (o_1, o_2, .., o_T)$，计算在模型$\lambda$下观测序列$O$，出现的概率 $P(O|\lambda)$?

# 问题1概率计算

给定模型$\lambda=(A,B,\pi)$，观察序列$O=(o_1,o_2,..,o_T)$,计算在模型$\lambda$下观测序列$O$，出现的概率 $P(O|\lambda)$？

- 法1：直接计算：列举所有可能的长度为$T$的状态序列，求各个状态序列$Q$和观测序列$O$的联合概率，然后对所有可能的状态序列求和。
- 给定一个固定的状态序列$\mathrm{Q}=(q_1,q_2,\ldots,q_T)$, 观测序列$O=(o_1,o_2,..,o_T)$的概率 $P(O|\mathrm{Q},\lambda)=\prod_{t=1}^{T}P(o_t|q_t,\lambda)=b_{q_1}(o_1)b_{q_2}(o_2)\cdots b_{q_T}(o_T)$

其中 $b_{q_t}(o_t)$表示在$q_t$状态下观测到$o_t$的概率

- 状态序列Q的概率$P(Q|\lambda)=\pi_{q_1}a_{q_1q_2}\ldots a_{q_{T-1}q_T}$
- O和Q同时出现的联合概率$P(O,Q|\lambda)=P(O|Q,\lambda)P(Q|\lambda)$
- 对所有的状态序列Q求和，得到序列O的概率：

$$P(O|\lambda)=\sum_{all\ Q}P(O,Q|\lambda)=\sum_{all\ Q}P(O|Q,\lambda)P(Q|\lambda)$$

# 法1:直接计算不可取。

$$P(O|\lambda) = \sum_{all\ Q} P(O,Q|\lambda) = \sum_{all\ Q} P(O|Q,\lambda)P(Q|\lambda)$$

$$= \sum_{q_1,q_2,\ldots,q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \ldots a_{q_{T-1} q_T} b_{q_\mathrm{T}}(o_T)$$

约 $2\mathrm{T} * N^T$计算量。 若N=5，T=100， $2 * 100 * 5^{100} \approx 10^{72}$

# 问题1：法2 前向法

- 定义前向变量：$\alpha_t(i)$表示到时刻$t$部分观测序列为$o_1, o_2, \cdots o_t$且状态为$q_t$的概率。 $\alpha_t(i) = P(\boldsymbol{o}_1, \boldsymbol{o}_2, \cdots \boldsymbol{o}_t, q_t = \mathrm{S}_i|\lambda) \qquad 1 \le t \le T$
- **算法2：前向法**：
- 输入：$\lambda, O$；输出：$\mathrm{P(O|\lambda)}$
  - 初始化：$\alpha_1(i) = \pi_i b_i(\boldsymbol{o}_1) \qquad \boldsymbol{i = 1, 2, \ldots, N}$
  - 递归：$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i)\, a_{ij}\right] b_j(\boldsymbol{o}_{t+1}) \qquad \boldsymbol{t = 1, 2}, \ldots, T-1, 1 \le j \le N$$
  - 终结：$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

# 解决问题1 前向法

$b_{q_t}(\boldsymbol{o}_t)$表示在$q_t$状态下观测到$O_t$的概率

$$\alpha_{t+1}(j) = P(\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_t, \boldsymbol{o}_{t+1}, \boldsymbol{q}_{t+1} = \boldsymbol{S}_j|\boldsymbol{\lambda})$$

$$= \sum_{i=1}^{N} P(\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_t, \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{o}_{t+1}, \boldsymbol{q}_{t+1} = \boldsymbol{S}_j|\boldsymbol{\lambda})$$

$$= \sum_{i}^{N} P(\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_t, \boldsymbol{q}_t = \boldsymbol{S}_i|\boldsymbol{\lambda}) * P(\boldsymbol{o}_{t+1}, \boldsymbol{q}_{t+1} = \boldsymbol{S}_j|\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_t, \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda})$$

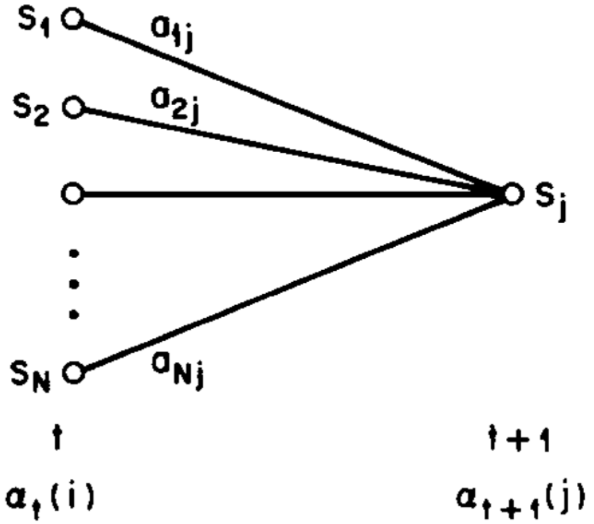
$$= \sum_{i}^{N} P(\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_t, \boldsymbol{q}_t = \boldsymbol{S}_i|\boldsymbol{\lambda}) * P(\boldsymbol{o}_{t+1}, \boldsymbol{q}_{t+1} = \boldsymbol{S}_j|\boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda})$$

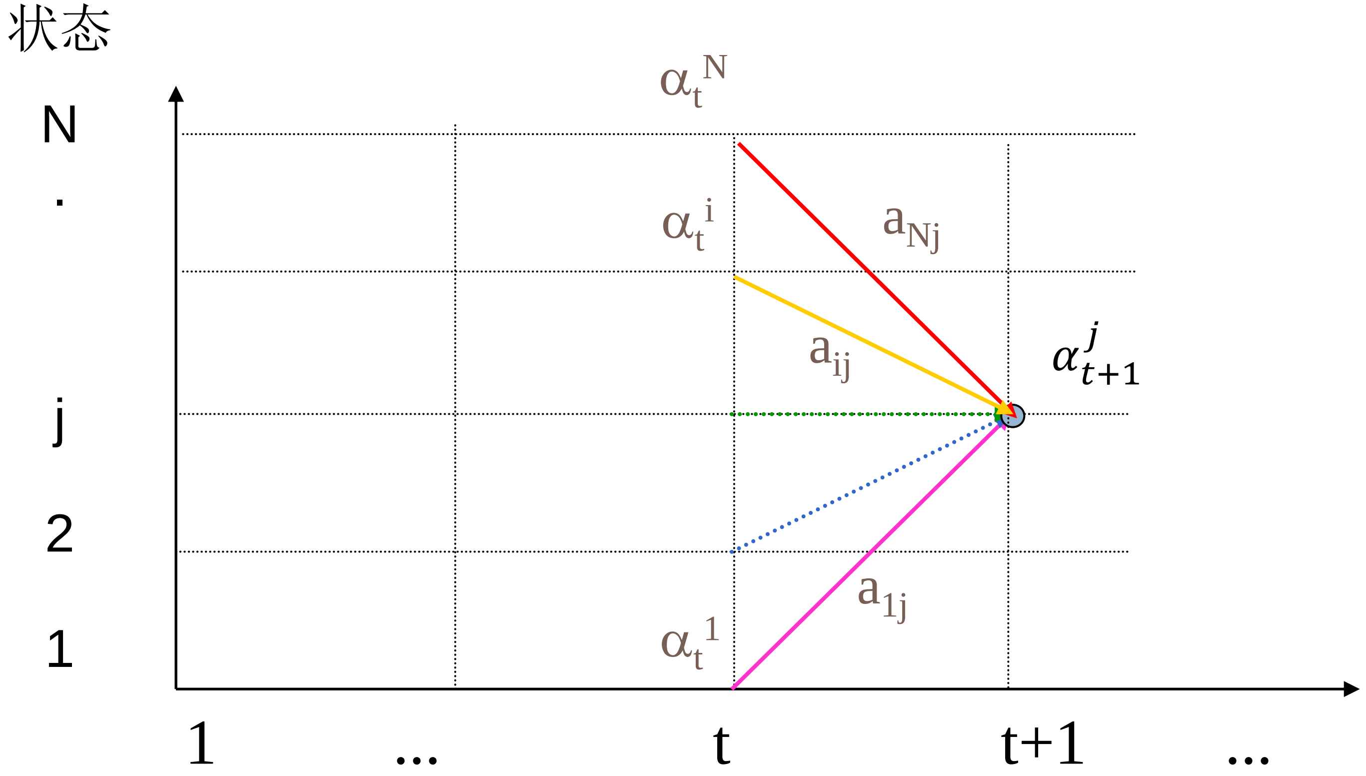
$$= \sum_{i}^{N} P(\boldsymbol{o}_1, \boldsymbol{o}_2, \dots, \boldsymbol{o}_t, \boldsymbol{q}_t = \boldsymbol{S}_i|\boldsymbol{\lambda}) * P\left(\boldsymbol{q}_{t+1} = \boldsymbol{S}_j|\boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda}\right) * P(\boldsymbol{o}_{t+1}|\boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda})$$

$$= \left[\sum_{i}^{N} \alpha_t(\boldsymbol{i}) a_{ij}\right] b_j(\boldsymbol{o}_{t+1})$$

灰色部分，表示可省略，即$P(\boldsymbol{o}_{t+1}|\boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda}) = P(\boldsymbol{o}_{t+1}|\boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{\lambda})$

# 前向法示意图




计算$\alpha_t(j), t = 1,2 \ldots, T, j = 1 \ldots, N.$，计算量 $NNT + N \sim O(\mathrm{T}N^2)$。 N=5, M=100，=> 计算量3000

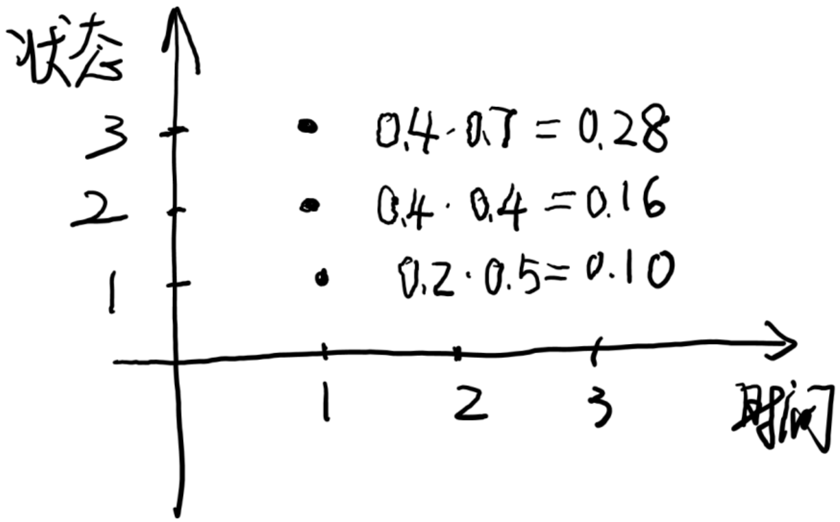
**例 10.2** 考虑盒子和球模型 $\lambda=(A,B,\pi)$，状态集合 $Q=\{1,2,3\}$，观测集合 $V=$ {红，白}，

$$A=\begin{bmatrix}0.5 & 0.2 & 0.3\\0.3 & 0.5 & 0.2\\0.2 & 0.3 & 0.5\end{bmatrix},\quad B=\begin{bmatrix}0.5 & 0.5\\0.4 & 0.6\\0.7 & 0.3\end{bmatrix},\quad \pi=(0.2,0.4,0.4)^{\mathrm{T}}$$

设 $T=3$，$O=$(红, 白, 红)，试用前向算法计算 $P(O|\lambda)$.

（1）计算初值

$$\alpha_1(1)=\pi_1 b_1(o_1)=0.10$$
$$\alpha_1(2)=\pi_2 b_2(o_1)=0.16$$
$$\alpha_1(3)=\pi_3 b_3(o_1)=0.28$$



**例 10.2** 考虑盒子和球模型 $\lambda=(A,B,\pi)$，状态集合 $Q=\{1,2,3\}$，观测集合 $V=\{红，白\}$，

$$A=\begin{bmatrix}0.5 & 0.2 & 0.3\\0.3 & 0.5 & 0.2\\0.2 & 0.3 & 0.5\end{bmatrix},\quad B=\begin{bmatrix}0.5 & 0.5\\0.4 & 0.6\\0.7 & 0.3\end{bmatrix},\quad \pi=(0.2,0.4,0.4)^{\mathrm{T}}$$

设 $T=3$， $O=(红, 白, 红)$，试用前向算法计算 $P(O\mid\lambda)$．

（2）递推计算

$$\alpha_2(1)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i1}\right]b_1(o_2)=0.154\times0.5=0.077$$

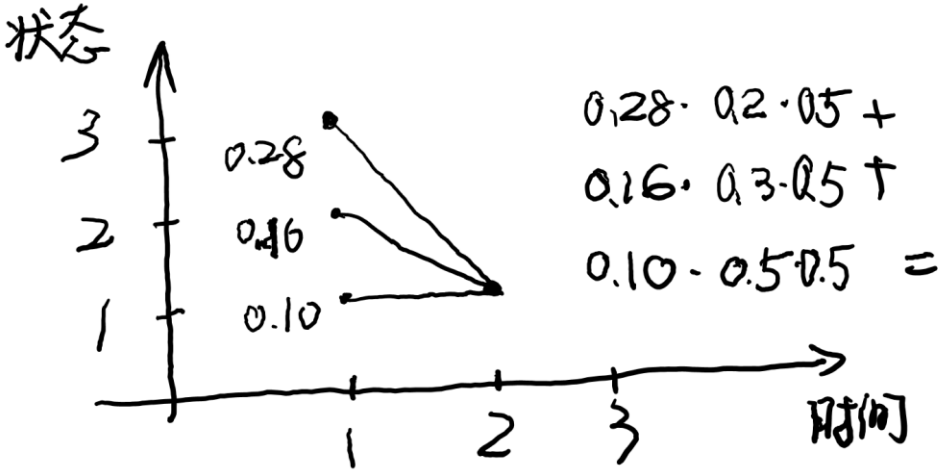
$$\alpha_2(2)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i2}\right]b_2(o_2)=0.184\times0.6=0.1104$$

$$\alpha_2(3)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i3}\right]b_3(o_2)=0.202\times0.3=0.0606$$

$$\alpha_3(1)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i1}\right]b_1(o_3)=0.04187$$

$$\alpha_3(2)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i2}\right]b_2(o_3)=0.03551$$

$$\alpha_3(3)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i3}\right]b_3(o_3)=0.05284$$




$0.28\cdot0.2\cdot0.5+$
$0.16\cdot0.3\cdot0.5+$
$0.10\cdot0.5\cdot0.5=0.077$

**例 10.2** 考虑盒子和球模型 $\lambda=(A,B,\pi)$，状态集合 $Q=\{1,2,3\}$，观测集合 $V=$ {红，白}，

$$A=\begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}, \quad B=\begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix}, \quad \pi=(0.2,0.4,0.4)^{\mathrm{T}}$$

设 $T=3$，$O=$(红, 白, 红)，试用前向算法计算 $P(O|\lambda)$.

（3）终止

$$P(O|\lambda)=\sum_{i=1}^{3}\alpha_3(i)=0.13022$$

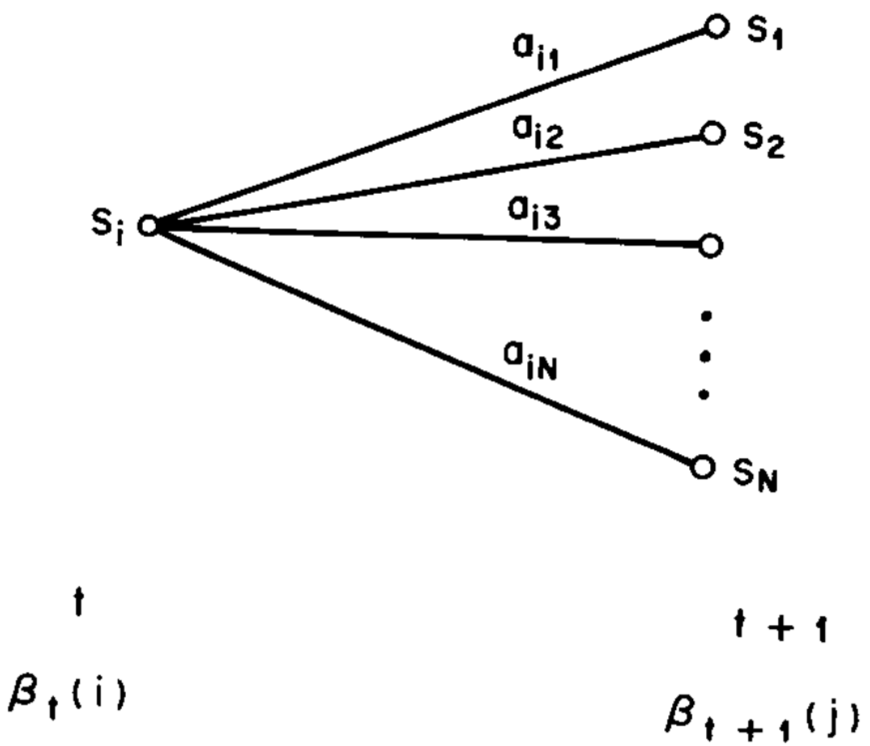
# 问题1：法3 后向法

- 与前向法类似，定义后向变量,表示从$t+1$时刻到结束，在时间$t$时刻，给定状态$S_i$，部分观测序列的概率。
- **算法3：向后算法：** $\beta_t(i)=P(\boldsymbol{o}_{t+1},\boldsymbol{o}_{t+2},\cdots\boldsymbol{o}_T|\lambda,q_t=\mathrm{S}_i)\qquad 1\le t\le T-1$
- 输入：$\lambda,O$; 输出：$\mathrm{P(O|}\lambda)$
  - 初始化：
    $$\beta_T(i)=1\qquad \boldsymbol{i=1,2,\ldots N}$$
  - 递归：
    $$\beta_t(i)=\sum_{j=1}^{N}a_{ij}b_j(\boldsymbol{o}_{t+1})\beta_{t+1}(j),\quad t=T-1,T-2,\ldots,1,1\le i\le N$$
  - 终结：
    $$P(O|\lambda)=\sum_{i=1}^{N}\pi_i\beta_1(i)\,b_i(\boldsymbol{o}_1)$$

# 问题1 后向法

$$\beta_t(i) = P(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \cdots \boldsymbol{o}_T | \lambda, q_t = S_i)$$
$$\boldsymbol{\beta_{t+1}(j)} = P(\boldsymbol{o}_{t+2}, \boldsymbol{o}_{t+3}, \cdots \boldsymbol{o}_T | \lambda, q_{\boldsymbol{t+1}} = S_{\boldsymbol{j}})$$

$$\beta_{\boldsymbol{t}}(\boldsymbol{i}) \overset{def}{=} P(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \cdots \boldsymbol{o}_T | q_t = S_i, \boldsymbol{\lambda})$$
$$= \sum_{j=1}^{N} P(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \cdots \boldsymbol{o}_T, \boldsymbol{q}_{t+1} = \boldsymbol{S}_j | \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda})$$
$$= \sum_{j}^{N} P(q_{t+1} = \boldsymbol{S}_j | \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda}) * P\left(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \cdots \boldsymbol{o}_T \middle| \boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{q}_t = \boldsymbol{S}_i, \boldsymbol{\lambda}\right)$$
$$= \sum_{j}^{N} \boldsymbol{a_{ij}} * P(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \cdots \boldsymbol{o}_T | \boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{\lambda})$$
$$= \sum_{j}^{N} \boldsymbol{a_{ij}} * \boldsymbol{P}\left(\boldsymbol{o}_{t+1} \middle| \boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{\lambda}\right) * P(\boldsymbol{o}_{t+2}, \cdots \boldsymbol{o}_T | \boldsymbol{o}_{t+1}, \boldsymbol{q}_{t+1} = \boldsymbol{S}_j, \boldsymbol{\lambda})$$
$$= \sum_{j} a_{ij} b_j(\boldsymbol{o}_{t+1}) \beta_{\boldsymbol{t+1}}(\boldsymbol{j})$$



$a_{i1}$
$S_1$
$a_{i2}$
$S_2$
$S_i$
$a_{i3}$
$a_{iN}$
$S_N$
$t$
$t+1$
$\beta_t(i)$
$\beta_{t+1}(j)$

# 前向后向算法

$$\mathrm{P}(\mathrm{O}|\lambda) = \sum_{i=1}^{N} P(o_1, \dots, o_T, q_t = S_i|\lambda)$$

$$= \sum_{i=1}^{N} P(o_1, \dots, o_t, q_t = S_i, o_{t+1}, \dots, o_T|\lambda) =$$

$$\sum_{i=1}^{N} P(o_1, \dots, o_t, q_t = S_i|\lambda) P(o_{t+1}, \dots, o_T|o_1, \dots, o_t, q_t = S_i, \lambda) =$$

$$\sum_{i=1}^{N} P(o_1, \dots, o_t, q_t = S_i|\lambda) P(o_{t+1}, \dots, o_T|q_t = S_i, \lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i)$$

## 问题3：预测问题或解码问题：

已知模型 $\lambda=(A,B,\pi)$ 和观测序列 $O=(o_1,o_2,..,o_T)$， 求最有可能对应的状态序列 $Q=(q_1,q_2,\dots,q_T)$。即，在该状态序列下，观测序列的条件概率 $P(O|Q)$ 最大?

# 问题3： Viterbi算法(动态规划)

- 目的：给定观察序列$O=(o_1,\ldots,o_T)$找到最合适的单状态序列$Q=(q_1,\ldots,q_T)$，使得Q能够最为合理的解释观察序列O。
- 首先定义目标函数：

$$\delta_t(i)=\max_{q_1,q_2,\ldots q_{t-1}} P[q_1q_2\ldots q_{t-1},q_t=S_i,\boldsymbol{o}_1,\boldsymbol{o}_2,\ldots\boldsymbol{o}_t,|\lambda]$$

即，$\delta_t(i)$是在$t$时刻，状态为$S_i$，所有单个路径中概率最大值。

通过递归可定义：

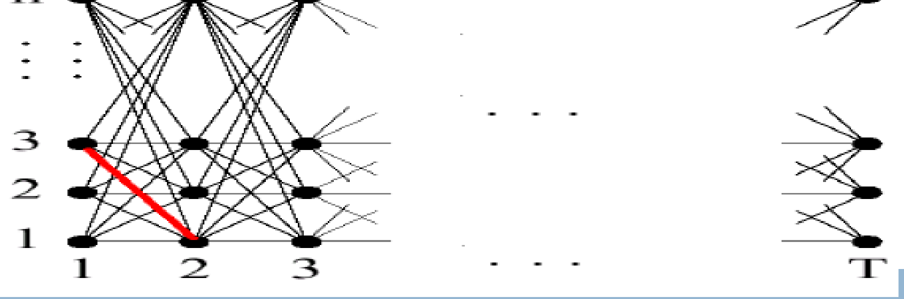
$$\delta_{t+1}(j)=\max_{q_1,q_2,\ldots q_{t-1}} P\left[q_1q_2\ldots q_{t-1},q_t,q_{t+1}=S_j,o_1,o_2,\ldots o_t,o_{t+1}\middle|\lambda\right]$$

$$=\left[\max_i \delta_t(i)a_{ij}\right]b_j(o_{t+1})$$

我们需要保存上述状态路径，定义

$$\boldsymbol{\psi_t(j)}=\boldsymbol{argmax_i}\left[\delta_{t-1}(i)a_{ij}\right]$$

# Viterbi算法(续)




## 算法4： Viterbi算法

- 输入：模型$\lambda$，观测$O$；输出最有路径$\mathrm{Q} = (\mathrm{q}_1^*, \ldots \mathrm{q}_\mathrm{T}^*)$
- 初始化：

$$\delta_1(i) = \pi_i b_i(\boldsymbol{o}_1), \qquad 1 \le \mathrm{i} \le \mathrm{N}$$

$$\boldsymbol{\psi}_1(i) = 0, \qquad 1 \le \mathrm{i} \le \mathrm{N}$$

- 递归：

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(\boldsymbol{o}_t), \qquad 2 \le t \le T, 1 \le j \le N$$

$$\boldsymbol{\psi}_t(j) = \underset{1 \le i \le N}{\operatorname{argmax}} [\delta_{t-1}(i) a_{ij}], \qquad 2 \le t \le T, 1 \le j \le N$$

- 终结：

$$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$

$$q_T^* = \underset{1 \le i \le N}{\operatorname{argmax}} [\delta_T(i)]$$

- 状态序列回溯：

$$q_t^* = \boldsymbol{\psi}_{t+1}(q_{t+1}^*), \qquad t = T-1, T-2, \ldots, 1$$

**例 10.3** 例 10.2 的模型 $\lambda=(A,B,\pi)$，

$$A=\begin{bmatrix}0.5 & 0.2 & 0.3\\0.3 & 0.5 & 0.2\\0.2 & 0.3 & 0.5\end{bmatrix},\quad B=\begin{bmatrix}0.5 & 0.5\\0.4 & 0.6\\0.7 & 0.3\end{bmatrix},\quad \pi=(0.2,0.4,0.4)^{\mathrm{T}}$$

已知观测序列 $O=$(红, 白, 红)，试求最优状态序列，即最优路径 $I^*=(i_1^*,i_2^*,i_3^*)$．

**解** 如图 10.4 所示，要在所有可能的路径中选择一条最优路径，按照以下步骤处理：

（1）初始化．在 $t=1$ 时，对每一个状态 $i$，$i=1,2,3$，求状态为 $i$ 观测 $o_1$ 为红的概率，记此概率为 $\delta_1(i)$，则

$$\delta_1(i)=\pi_i b_i(o_1)=\pi_i b_i(\text{红}),\quad i=1,2,3$$

代入实际数据 0.2*0.5 0.4*0.4 0.4*0.7

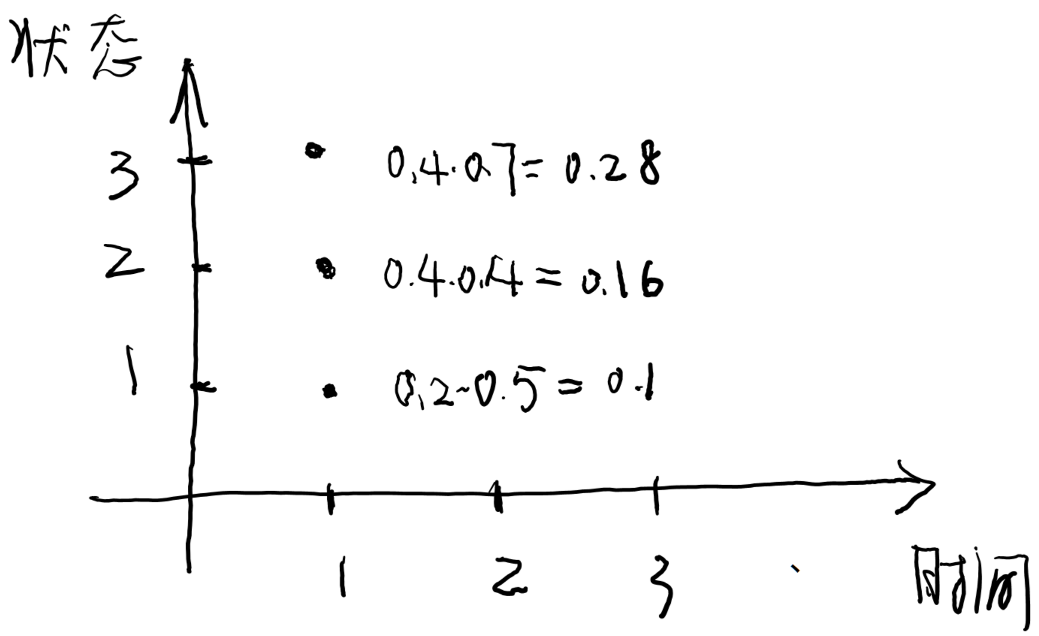
$$\delta_1(1)=0.10,\quad \delta_1(2)=0.16,\quad \delta_1(3)=0.28$$

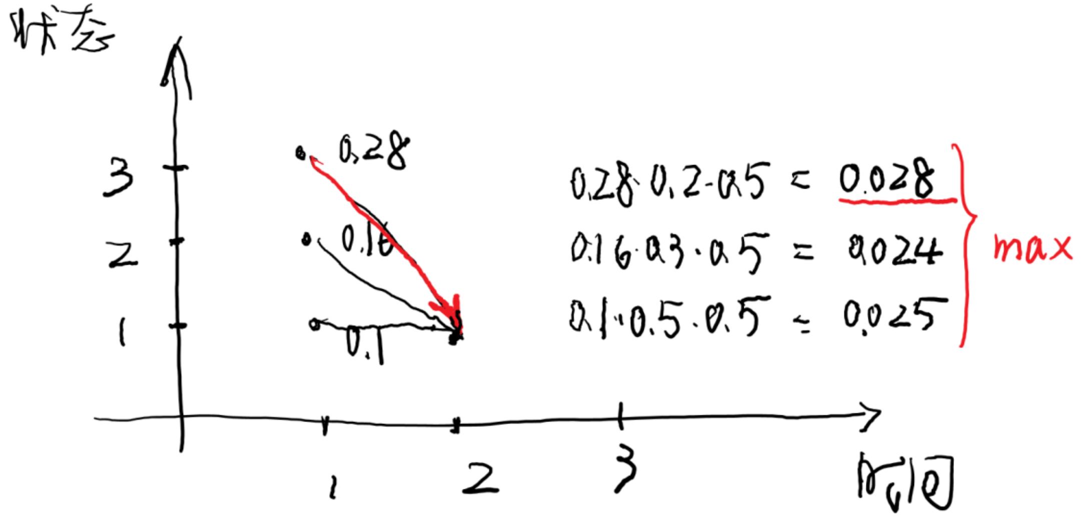
记 $\psi_1(i)=0$，$i=1,2,3$．

**例 10.3** 例 10.2 的模型 $\lambda=(A,B,\pi)$，

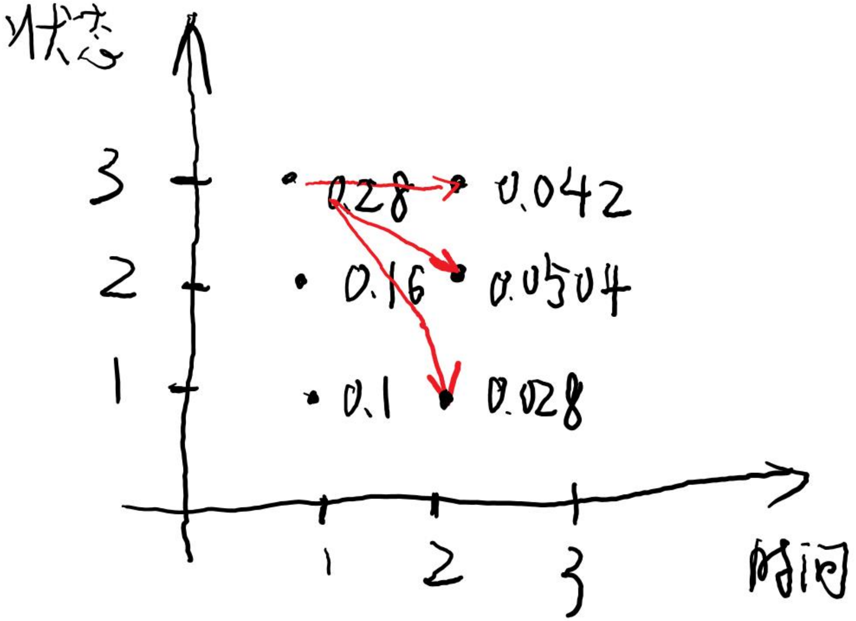
$$A=\begin{bmatrix}0.5 & 0.2 & 0.3\\0.3 & 0.5 & 0.2\\0.2 & 0.3 & 0.5\end{bmatrix},\quad B=\begin{bmatrix}0.5 & 0.5\\0.4 & 0.6\\0.7 & 0.3\end{bmatrix},\quad \pi=(0.2,0.4,0.4)^{\mathrm{T}}$$

已知观测序列 $O=$(红, 白, 红)，试求最优状态序列，即最优路径 $I^*=(i_1^*,i_2^*,i_3^*)$ .

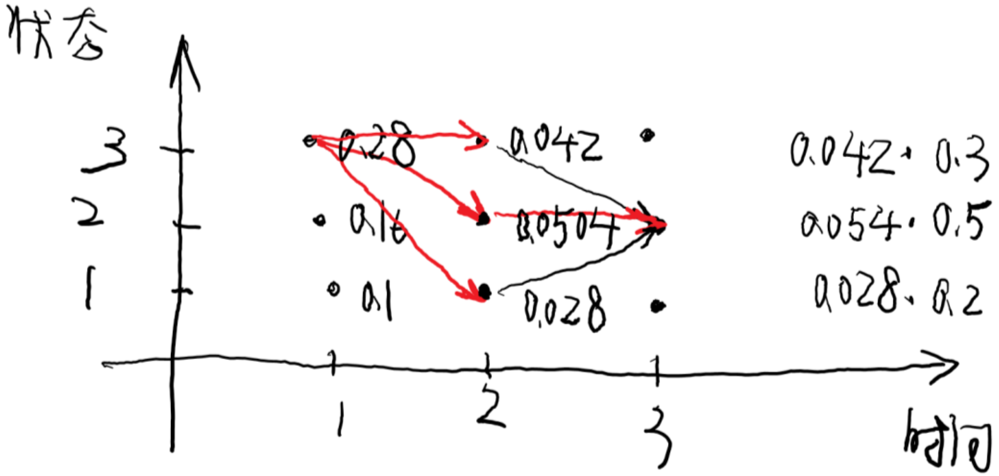




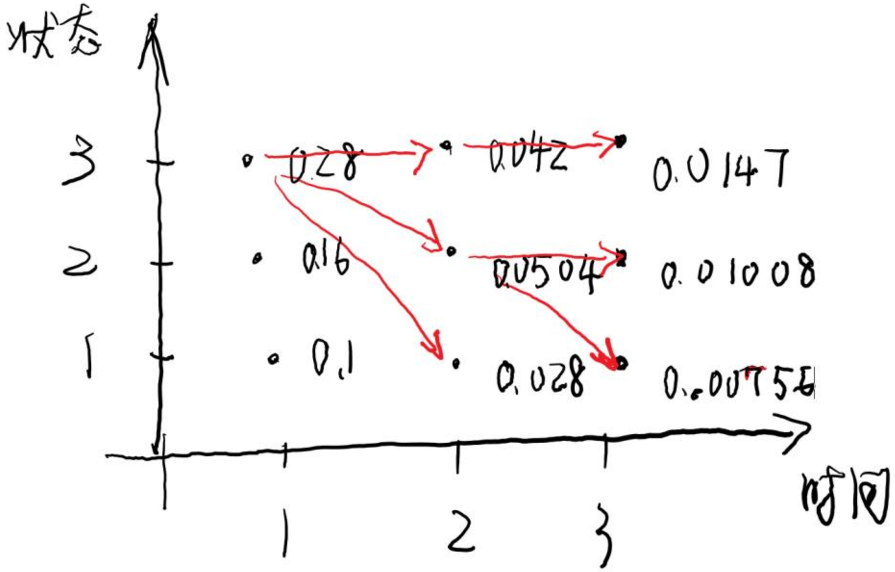
状态
3
2
1
0.28
0.16
0.1
1
2
3
时间
0.28·0.2·0.5 = 0.028
0.16·0.3·0.5 = 0.024
0.1·0.5·0.5 = 0.025
max

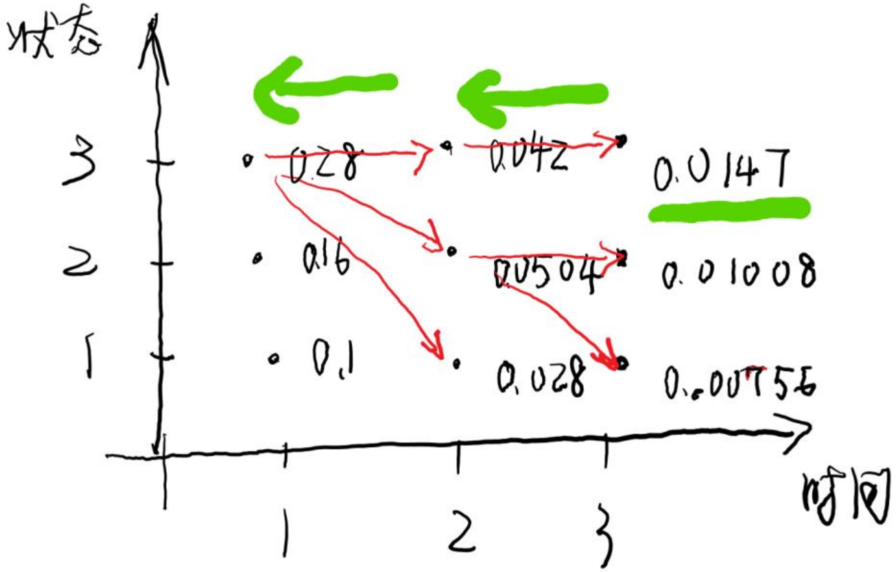



状态
3
2
1
0.28
0.16
0.1
0.042
0.0504
0.028
1
2
3
时间




$$\max\left\{\begin{aligned} &0.042\cdot 0.3\cdot 0.4 = 0.00504\\ &0.054\cdot 0.5\cdot 0.4 = \underline{0.01008}\\ &0.028\cdot 0.2\cdot 0.4 = 0.00224 \end{aligned}\right.$$



（2）在 $t=2$ 时，对每个状态 $i$， $i=1,2,3$，求在 $t=1$ 时状态为 $j$ 观测为红并在 $t=2$ 时状态为 $i$ 观测 $o_2$ 为白的路径的最大概率，记此最大概率为 $\delta_2(i)$，则

$$\delta_2(i)=\max_{1\leqslant j\leqslant 3}[\delta_1(j)a_{ji}]b_i(o_2)$$

同时，对每个状态 $i$， $i=1,2,3$，记录概率最大路径的前一个状态 $j$：

$$\psi_2(i)=\arg\max_{1\leqslant j\leqslant 3}[\delta_1(j)a_{ji}], \qquad i=1,2,3$$

计算：

$$\begin{aligned}\delta_2(1)&=\max_{1\leqslant j\leqslant 3}[\delta_1(j)a_{j1}]b_1(o_2)\\&=\max_j\{0.10\times 0.5,0.16\times 0.3,0.28\times 0.2\}\times 0.5\\&=0.028\end{aligned}$$

$$\psi_2(1)=3$$

$$\delta_2(2)=0.0504,\quad \psi_2(2)=3$$

$$\delta_2(3)=0.042,\quad \psi_2(3)=3$$

同样，在 $t=3$ 时，

$$\delta_3(i)=\max_{1\leqslant j\leqslant 3}[\delta_2(j)a_{ji}]b_i(o_3)$$

$$\psi_3(i)=\arg\max_{1\leqslant j\leqslant 3}[\delta_2(j)a_{ji}]$$

$$\delta_3(1)=0.00756,\quad \psi_3(1)=2$$

$$\delta_3(2)=0.01008,\quad \psi_3(2)=2$$

$$\delta_3(3)=0.0147,\quad \psi_3(3)=3$$

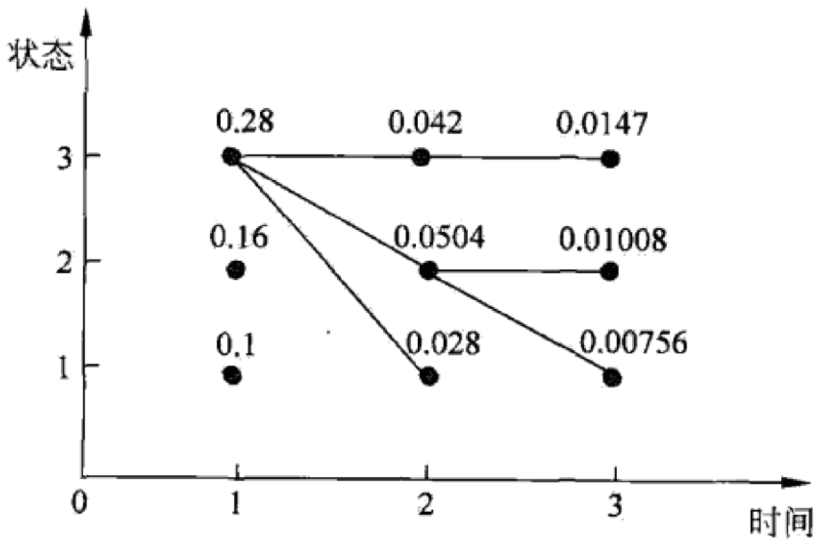



图 10.4　求最优路径

（3）以 $P^*$ 表示最优路径的概率，则

$$P^* = \max_{1\leqslant i\leqslant 3} \delta_3(i) = 0.0147$$

最优路径的终点是 $i_3^*$：

$$i_3^* = \arg\max_i [\delta_3(i)] = 3$$

（4）由最优路径的终点 $i_3^*$，逆向找到 $i_2^*, i_1^*$：

$$\text{在 } t=2 \text{ 时，} \quad i_2^* = \psi_3(i_3^*) = \psi_3(3) = 3$$

$$\text{在 } t=1 \text{ 时，} \quad i_1^* = \psi_2(i_2^*) = \psi_2(3) = 3$$

于是求得最优路径，即最优状态序列 $I^* = (i_1^*, i_2^*, i_3^*) = (3,3,3)$ .

# 解决问题2：学习问题(最困难)

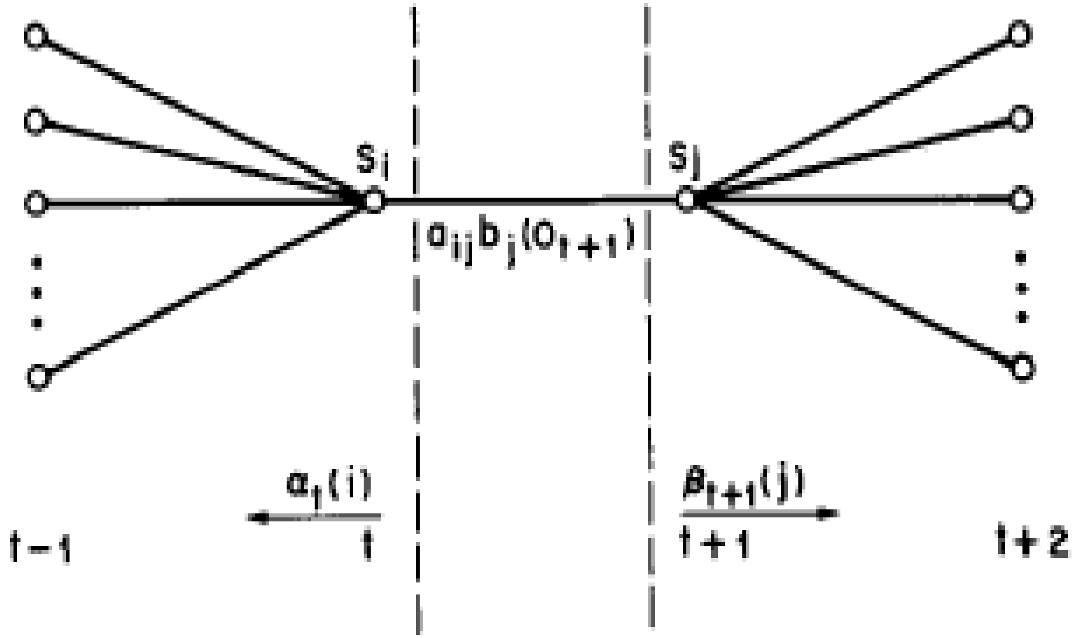
已知观测序列$O=(o_1,o_2,..,o_T)$，估计模型$\lambda=(A,B,\pi)$参数，使得在该参数下，观测序列概率$P(O|\lambda)$最大。即，用极大似然估计方法估计参数。

第一种情况，我们已知观测序列O，和状态序列(隐藏序列）S。

$$O_1 = (o_1, \dots o_T), S_1 = (q_1, \dots, q_T)$$

# 引入两个变量$\xi_t(i,j)$

- 定义$\xi_t(i,j)$表示给定观测序列$O$，$t$时刻状态为$S_i$，$t+1$时刻状态为$S_j$的概率。
- $\xi_t(i,j) = P\left(q_t = S_i, q_{t+1} = S_j \mid O, \lambda\right)$




$$\xi_t(\boldsymbol{i},\boldsymbol{j}) = \frac{P\left(q_t = S_i, q_{t+1} = S_j, O|\lambda\right)}{P(O \mid \lambda)}$$

$$= \frac{\alpha_t(\boldsymbol{i})a_{ij}b_j(\boldsymbol{o}_{t+1})\beta_{t+1}(\boldsymbol{j})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(\boldsymbol{i})a_{ij}b_j(\boldsymbol{o}_{t+1})\beta_{t+1}(\boldsymbol{j})}$$

# 引入两个变量$\gamma_t(i)$

$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda)$

- 定义$\gamma_t(i)$ 为给定观察序列$O$和模型$\lambda$之后，在时间$t$时刻，状态为$S_i$的概率。
- $\gamma_t(i) = P(q_t = S_i \mid O, \lambda) = \frac{P(O, q_t = S_i|\lambda)}{P(O|\lambda)} = \sum_{j=1}^{N} \xi_t(i,j)$
- $\gamma_t(i)$和前向概率，后向概率的关系
  - $\alpha_t(i)$在$t$时刻，状态为$S_i$的观测序列为$o_1, \ldots, o_t$ 的概率
  - $\beta_t(i)$在$t$时刻，状态为$S_i$的观测序列为$o_{t+1}, \ldots, o_T$ 的概率
  - $\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}$

# 将$\xi_t(i,j)$，$\gamma_t(i)$ 求和可得到有用的期望值。

- 在观测序列O下，状态$S_i$出现的期望数。
- $\sum_{t=1}^{T} \gamma_t(i)$

- 在观测序列O下，状态$S_i$ 转移的期望数。
- $\sum_{t=1}^{T-1} \gamma_t(i)$

- 在观测序列O下，状态$S_i$转移到$S_j$的期望数。
- $\sum_{t=1}^{T-1} \xi_t(i,j)$

# Baum-Welch算法

- 基于以上定义，可以得到HMM的参数估计：

**初始状态$\pi_i$ 的估计 $\hat{\pi}_i$：** 为样本中初始状态为$S_i$ 频率 (或 t＝1时，$\gamma_1(i)$ )

$$\text{状态转移概率的估计}\hat{a}_{ij} = \frac{\text{expected number of transitions from S\_i to S\_j}}{\text{expected number of transitions from S\_i}} =$$

$$\frac{\text{样本中时刻t处于状态Si，时刻t+1转移到SJ的频数A\_ij}}{\sum_{j=1,..N} A_{ij}} =$$

$$\frac{\sum_{t=1...T-1} \xi_t(i,j)}{\sum_{t=1,...,T-1} \sum_{j=1,...N} \xi_t(i,j)} = \frac{\sum_{t=1...T-1} \xi_t(i,j)}{\sum_{t=1,...,T-1} \gamma_t(i)}$$

$$\text{观测概率的估计}\hat{b}_j(k) = \frac{\text{expected number of times in state } j \text{ and observing symbol } vk}{\text{expected number of times in state } j}$$

$$= \frac{\sum_{t=1,...,T,s.t.O_t=v_k} \gamma_t(j)}{\sum_{t=1,...,T} \gamma_t(j)}$$

- 第二种情况，只有观测序列O

# Baum-Welch 算法(EM算法原理)

- 给定HMM模型$\lambda=(A,B,\pi)$，观测数据为O，状态序列为Q，那么HMM可看做一个含有隐变量Q的概率模型。
- $P(O|\lambda)=\sum_t P(O|Q,\lambda)P(Q|\lambda)$
- 它的参数可以通过EM算法来求解。
  - EM算法E步：求联合分布$P(O,Q|\lambda)$基于条件概率$P(Q|O,\lambda')$的期望。其中$\lambda'$是当前模型的参数。然后M步最大化这个期望，得到更新的模型参数$\lambda$
  - $Q(\lambda,\lambda')=\sum_Q P(Q|O,\lambda')\log P(O,Q|\lambda)$,其中$\lambda'$是 HMM参数的当前估计。$\lambda$要极大化HMM的参数。
  - $P(O,Q\mid\lambda)=\pi_{q_1}b_{q_1}(o_1)a_{q_1q_2}b_{q_2}(o_2)\ldots a_{q_{T-1}q_T}b_{q_T}(o_T)$

- $P(O,Q \mid \lambda) = \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \ldots a_{q_{T-1} q_T} b_{q_T}(o_T)$

- $Q(\lambda,\lambda') = \sum_Q P(Q|O,\lambda')\, logP(O,Q|\lambda) = \sum_Q \frac{P(Q,O|\lambda')}{P(O|\lambda')} \log P(O,Q|\lambda) = \frac{\sum_Q P(Q,O|\lambda') logP(O,Q|\lambda)}{P(O|\lambda')} \sim \sum_Q P(Q,O|\lambda') \log P(O,Q|\lambda)$
- 优化$Q(\lambda,\lambda')$ 等价优化$\sum_Q P(Q,O|\lambda') \log P(O,Q|\lambda)$ （左式仍记为$Q(\lambda,\lambda')$）
- $Q(\lambda,\lambda') = \sum_Q \log \pi_{q_1} P(O,Q \mid \lambda') + \sum_Q \left(\sum_{t=1}^{T-1} \log a_{q_t q_{t+1}}\right) P(O,Q \mid \lambda') + \sum_Q \left(\sum_{t=1}^{T} \log b_{q_t}(o_t)\right) P(O,Q \mid \lambda')$
-

$$Q(\lambda,\lambda') = \sum_Q \log \pi_{q_1} P(O,Q \mid \lambda') + \sum_Q \left(\sum_{t=1}^{T-1} \log a_{q_t q_{t+1}}\right) P(O,Q \mid \lambda') + \sum_Q \left(\sum_{t=1}^{T} \log b_{q_t}(o_t)\right) P(O,Q \mid \lambda')$$

- M步，极大化Q函数，求模型参数$\lambda$。由于参数单独出现在三项中（$\pi, A, B$），可分别求最大化。
- 第一项：$\sum_Q \log \pi_{q_1} P(O,Q \mid \lambda') = \sum_{i=1,\ldots,N} \log \pi_i P(O, q_1 = S_i \mid \lambda')$
- 拉格朗日函数

$$\frac{\partial}{\partial \pi_i}\left(\sum_{i=1}^{N} \log\pi_i \, P(O, q_1 = S_i \mid \lambda') + \gamma\left(\sum_{i=1}^{N} \pi_i - 1\right)\right) = 0$$

得$\frac{1}{\pi_i} P(O, q_1 = S_i \mid \lambda') + \gamma = 0 \rightarrow P(O, q_1 = S_i \mid \lambda') + \pi_i \gamma = 0$

两边对i求和 得 $\gamma = -P(O|\lambda') \rightarrow \pi_i = \frac{P(O, q_1 = S_i | \lambda')}{P(O|\lambda')}$

- 第二项： $\sum_Q \left(\sum_{t=1}^{T-1} \log a_{q_t q_{t+1}}\right) P(O, Q \mid \lambda') = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij}\, P(O, q_t = S_i, q_{t+1} = S_j \mid \lambda')$
- 拉格朗日函数求导
- $\frac{\partial}{\partial a_{ij}} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij}\, P(O, q_t = S_i, q_{t+1} = S_j \mid \lambda') + \gamma\left(\sum_{j=1}^{N} a_{ij} - 1\right)\right) = 0$
- $P\left(O, q_t = S_i, q_{t+1} = S_j \mid \lambda'\right) + a_{ij}\gamma = 0$，对$j$求和
- $\gamma = -P(O, q_t = S_i \mid \lambda')$
- $a_{ij} = \frac{\sum_{t=1}^{T-1} P\left(O, q_t = S_i, q_{t+1} = S_j \mid \lambda'\right)}{\sum_{t=1}^{T-1} P\left(O, q_t = S_i \mid \lambda'\right)}$

- 第三项：$\sum_Q \left(\sum_{t=1}^{T} \log b_{q_t}(o_t)\right) P(O,Q \mid \lambda') = \sum_{j=1}^{N} \sum_{t=1}^{T} \log b_j(o_t) P(O, q_t = S_j \mid \lambda')$
- 拉格朗日函数求导
- $\frac{\partial}{\partial b_j(k)} \left( \sum_{j=1}^{N} \sum_{t=1}^{T} \log b_j(o_t) P(O, q_t = S_j \mid \lambda') + \gamma\left(\sum_{k=1}^{M} b_j(o_t = v_k) - 1\right)\right) = 0$
- $\left(\sum_{t=1}^{T} \frac{1}{b_j(k)} P(O, q_t = S_j \mid \lambda') I(o_t = v_k)\right) + \gamma = 0$, 其中$I(o_t = v_k)$表示当$o_t = v_k$取1，否则取0.
- $\sum_{t=1}^{T} P(O, q_t = S_j \mid \lambda') I(o_t = v_k) + b_j(k)\gamma = 0$对$k$求和
- $\gamma = -\sum_{t=1}^{T} P(O, q_t = S_j \mid \lambda')$
- $b_j(k) = \frac{\sum_{t=1}^{T} P(O, q_t = S_j \mid \lambda') I(o_t = v_k)}{\sum_{t=1}^{T} P(O, q_t = S_j \mid \lambda')}$

$$\gamma_t(i) = P(q_t = S_i \mid O, \lambda) = \frac{P(O, q_t = S_i|\lambda)}{P(O|\lambda)}$$

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda) = \frac{P(\boldsymbol{O}, q_t = S_i, q_{t+1} = S_j|\lambda)}{P(\boldsymbol{O}|\lambda)}$$

- 使用$\gamma_i(t), \xi_t(i,j)$表示结果

$$\pi_i = \frac{P(O, q_1 = S_i \mid \lambda')}{P(O|\lambda')} = \boldsymbol{\gamma_i(1)}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, q_t = S_i, q_{t+1} = S_j|\lambda')}{\sum_{t=1}^{T-1} P(O, q_t = S_i|\lambda')} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$b_j(k) = \frac{\sum_{t=1}^{T} P(O, q_t = S_j \mid \lambda') I(o_t = v_k)}{\sum_{t=1}^{T} P(O, q_t = S_j \mid \lambda')} = \frac{\sum_{t=1, o_t = v_k}^{T} \boldsymbol{\gamma_t(j)}}{\sum_{t=1}^{T} \boldsymbol{\gamma_t(j)}}$$

- 和第一种情况比较

**初始状态$\boldsymbol{\pi_i}$ 的估计 $\hat{\pi}_i$：** $\gamma_1(i)$

**状态**转移概率的估计$\boldsymbol{\hat{a}_{ij}} =$

$$\frac{\sum_{t=1\ldots T-1} \xi_t(i,j)}{\sum_{t=1,\ldots,T-1} \gamma_t(i)}$$

**观测**概率的估计$\boldsymbol{\hat{b}_j(k)} = \frac{\sum_{t=1,\ldots,T, s.t. O_t = v_k} \gamma_t(j)}{\sum_{t=1,\ldots,T} \gamma_t(j)}$

- **算法5： Baum-Welch算法**
- 输入：观测数据O，输出模型参数$\lambda$
- 初始化： $n=0, a_{ij}^{(0)}, b_j(k)^{(0)}, \pi_i^{(0)}$, 模型$\lambda^{(0)}=(A^{(0)}, B^{(0)}, \pi^{(0)})$
- 递推，$n=1,2,\ldots,\ldots$ $\pi_i^{(n+1)}=\gamma_i(1)$

$$a_{ij}^{(n+1)}=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

$$b_j(k)^{(n+1)}=\frac{\sum_{t=1,o_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}$$

- 终止：得到模型：$\lambda^{(n+1)}=(A^{(n+1)}, B^{(n+1)}, \pi^{(n+1)})$

# HMM中的三个问题

| 问题 | 解决方法 | 本质 | |
|---|---|---|---|
| 识别问题 | 向前（前向）算法 | 递归计算 | |
| 学习问题 | Baum-Welch 算法 | EM算法 | |
| 解码问题 | Viterbi算法 | 动态规划算法 | |
| | | | |

# HMM应用

| 问题 | 观测值 | 隐藏状态 | | |
|---|---|---|---|---|
| 语音识别 | 语音信号 | 文字 | | |
| 机器中英文翻译 | 中文 | 英文 | | |
| 拼音输入法 | 拼音 | 中文候选文字 | | |

**https://zhuanlan.zhihu.com/p/111899116**