Hmm

hmm 三个问题

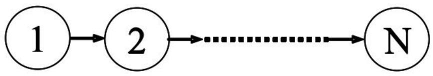
1. 给定参数 A B Pi，求一个观察到显状态序列的概率
2. 解码，给定参数 A B Pi，求观察序列背后的隐状态序列
3. 求参数 A B Pi　其中 A 是隐状态之间转移概率，B 是发射概率，隐状态发射到显状态到概率，Pi 是初始隐状态概率

hmm 的假设：

1. 一阶马尔科夫假设。隐状态只跟前一状态有关，跟其他状态无关
2. 观测状态只与当前的隐状态有关，跟其他状态无关

第一个问题：求观察到状态概率

全概率公式，会造成时间复杂度 O(n^t)问题。原因是每一个隐状态都会有产生特定显状态的概率，并且转移概率



$\lambda = (A, B, \pi)$

$\overline{O} = o_1 o_2 o_3 \ldots\ldots o_t \ldots\ldots o_T$ observation sequence

$\overline{q} = q_1 q_2 q_3 \ldots\ldots q_t \ldots\ldots q_T$ state sequence

- **Problem 1:** Given λ and $\overline{O}$,
  find $P(\overline{O}|\lambda)$=Prob[observing $\overline{O}$ given λ]
- **Direct Evaluation: considering all possible** state sequence $\overline{q}$

$$P(\overline{O}|\lambda) = \sum_{\text{all } \overline{q}} P(\overline{O}, \overline{q}|\lambda) = \sum_{\text{all } \overline{q}} P(\overline{O}|\overline{q}, \lambda) P(\overline{q}|\lambda)$$

$$P(\overline{O}|\overline{q}, \lambda)$$

$$\Uparrow$$

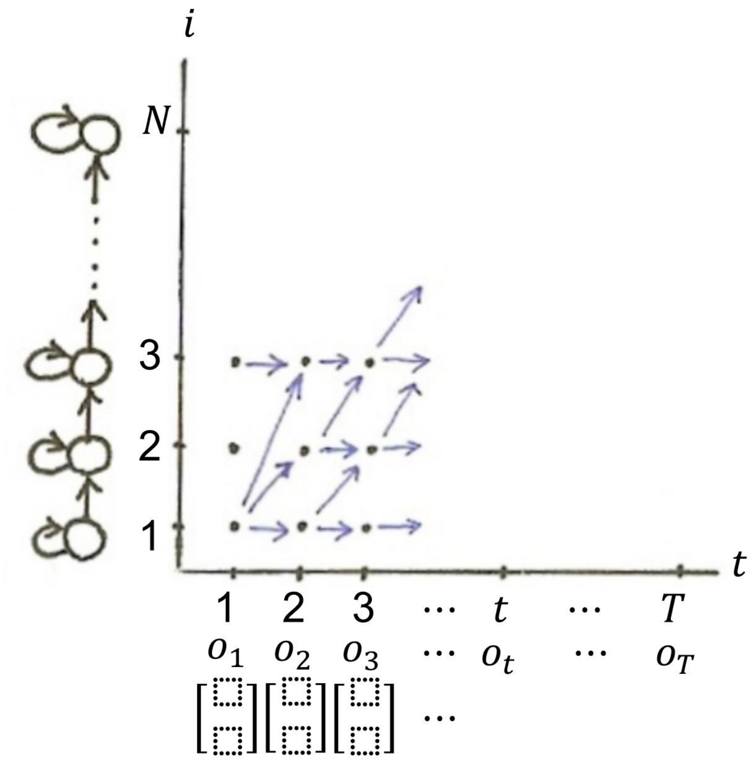
$$P(\overline{O}|\lambda) = \sum_{\text{all } \overline{q}} \Big([b_{q_1}(o_1) \bullet b_{q_2}(o_2) \bullet \ldots\ldots b_{q_T}(o_T)] \bullet [\pi_{q_1} \bullet a_{q_1 q_2} \bullet a_{q_2 q_3} \bullet \ldots\ldots a_{q_{T-1} q_T}] \Big)$$

$$\Downarrow$$

$$P(\overline{q}|\lambda)$$

total number of different $\overline{q}$: N^T

huge computation requirements



纵坐标代表隐状态，横坐标是时间轴，o1,o2...ot 指的是第 t 个时间点第观察状态。
将概率问题转化为路径问题。转化为路径之后就想到了二维表上第动态规划，可以优化那个高阶时间复杂度第计算问题。

Basic Problem 1 for HMM

- Forward Algorithm: defining a forward variable $\alpha_t(i)$

$$\alpha_t(i) = P(o_1 o_2 \ldots\ldots o_t\, ,\, q_t = i|\lambda)$$

$$= \text{Prob}[\text{observing } o_1 o_2 \ldots o_t\, , \text{ state i at time } t|\lambda]$$

- Initialization

$$\alpha_1(i) = \pi_i b_i(o_1)\ ,\quad 1 \le i \le N$$

- Induction

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}]\, b_j(o_{t+1})$$

$$1 \le j \le N$$

$$1 \le t \le T-1$$

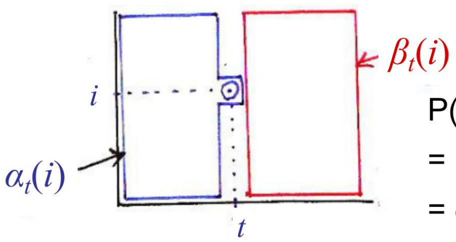
- Termination

$$P(\overline{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

See Fig. 6.5 of Rabiner and Juang

- All state sequences, regardless of how long previously, merge to the N state at each time instant t

理解 hmm 的第一个问题，只需要知道是利用了前向、后向算法，其中前向后向算法是利用动态规划来优化指数的大 o 阶问题即可，具体公式不需要记住，因为记住几天就会忘。
同时做了假设，就如下图所示：A 和 B 相互独立假设，意味着观测状态 t 之前和之后是相互独立的。这个假设是不合理的。但是这并不影响 hmm 是一个效果好的算法

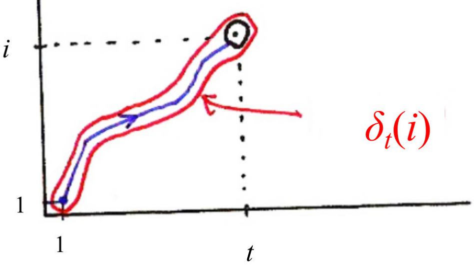


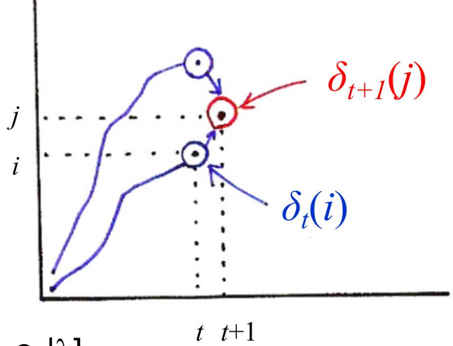
$$P(\bar{O}, q_t = i|\lambda)$$
$$= \text{Prob}\,[\text{observing } o_1, o_2, \ldots, o_t, \ldots, o_T, q_t = i|\lambda]$$
$$= \alpha_t(i)\beta_t(i)$$

$$A: (o_1\ o_2\ \cdots o_t|\lambda)$$
$$B: (o_{t+1},\ o_{t+2},\ \cdots o_T|\lambda)$$
$$C: (q_t = i|\lambda)$$
$$P(A,\ B,\ C) = P(A,\ C)\ P(B|A,\ C)$$
$$/\!/ \qquad\qquad /\!/ \qquad\qquad /\!/ \qquad\qquad (B \perp A)$$
$$P(\bar{O}, q_t = i|\lambda) \quad \alpha_t(i) \quad P(B|C)$$
$$/\!/$$
$$\beta_t(i)$$

第二个问题 解码问题
维特比算法。维特比算法其实就是动态规划算法。动态规划就是找第 t+1 项跟前面几项或者 1 项的关系，在 hmm 里面是跟前面第 t 项之间的关系。解码问题其实就是求当前观测序列条件下最可能的隐状态，用概率表达，其实就是最大联合概率密度下式中的 delta。
接下来就是求 delta 第 t+1 项和第 t 项之间的一种关系（动态规划的套路），然后求解初始值（动态规划的套路），然后就可以迭代或者递归求第 t 项的值。
1. 规律。转化为路径规划问题之后很容易发现规律
2. 初始值。因为参数都是已知的，所以 delta0(i)就是 pi，A 和 B 也都是已知的，所以可以用维特比来解码了

$$\delta_{t+1}(j) = \max_i\, [\delta_t(i)a_{ij}] \bullet b_j(o_{t+1})$$





$$\delta_t(i) = \max_{q_1,q_2,\ldots q_{t-1}} P[q_1,q_2,\ldots q_{t-1},\ q_t = i,\ o_1,o_2,\ldots,o_t\,|\lambda]$$

第三个问题：参数求解问题