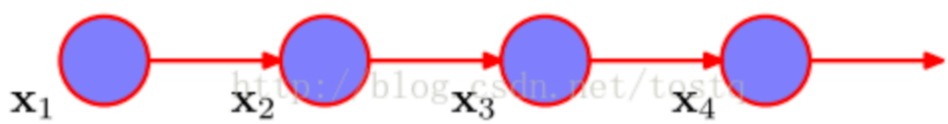
# 引入

首先，已知一组序列：

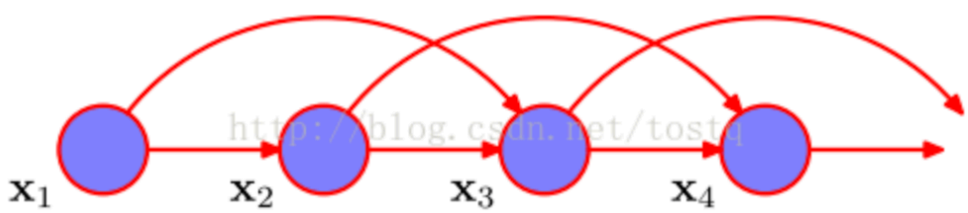

我们从这组序列中推导出产生这组序列的函数，假设函数参数为 ，其表示为

$$\theta = arg \max_{\theta} P(X|\theta)$$

即**使得序列X发生概率最大的函数参数**，要解决上式，最简单的考虑是将序列 的每个数据都视为独立的，比如建立一个神经网络。然后这种考虑会随着序列增长，而导致参数爆炸式增长。因此可以**假设当前序列数据只与其前一数据值相关**，**即所谓的一阶马尔科夫链**：




有一阶马尔科夫链，也会有二阶马尔科夫链（即当前数据值取决于其前两个数据值）




当前本文不对二阶马尔科夫链进行深入分析了，着重考虑一阶马尔科夫链，现在根据一阶马尔科夫链的假设，我们有：

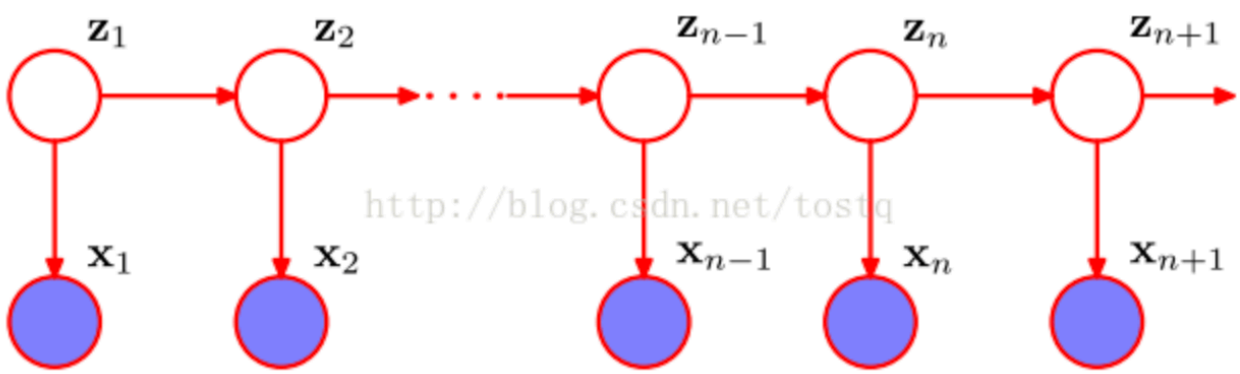
$$P(x_n|x_1,\cdots,x_{n-1}) = P(x_n|x_{n-1})$$

因此要解一阶马尔科夫链，其关键在于求数据（以下称观测值）之间转换函数，如果**假设转换函数**$P(x_n|x_{n-1})$**同序列中位置（时间）无关**，我们就能根据转换函数$P(x_n|x_{n-1})$而求出整个序列的概率：

$$P(x_1,\cdots,x_N) = P(x_1)\prod_{n=2}^{N} P(x_n|x_{n-1})$$

然而，如果观测值x的状态非常多（特别极端的情况是连续数据），转换函数$P(x_n|x_{n-1})$会变成一个非常大的矩阵，如果x的状态有K个，那么转换函数$P(x_n|x_{n-1})$就会是一个K*(K-1)个参数，而且对于连续变量观测值更是困难。

为了降低马尔科夫链的转换函数的参数量，我们引入了一个包含较少状态的隐状态值，将观测值的马尔科夫链转换为隐状态的马尔科夫链（即为隐马尔科夫链HMM）




其包含了一个**重要假设：当前观测值只由当前隐状态所决定**。这么做的一个重要好处是，隐状态值的状态远小于观测值的状态，因此隐藏状态的转换函数$P(z_n|z_{n-1})$ 的参数更少。

此时我们要决定的问题是：

$$\theta = arg\max_{\theta} P(X|\theta) = arg\max_{\theta} \sum_{Z} P(X,Z|\theta)$$



即**在所有可能隐藏状态序列情况下，求使得序列 发生概率最大的函数参数** 。

# 转移概率

含义
- $A_{jk} = p(z_{nk}=1 \mid z_{n-1,j}=1)$.

  $0 \le A_{jk} \le 1$ $\quad \sum_k A_{jk} = 1$

参数量
- 有 $K(K-1)$ 参数

计算式
- $p(z_n \mid z_{n-1}, A) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1,j} z_{nk}}$.

$p(z_1 \mid \pi)$
- 初始潜在变量 $z_1$,

  $p(z_1 \mid \pi) = \prod_{k=1}^{K} \pi_k^{z_{1k}}$.

  $\sum \pi_k = 1$ 表示初始为每个状态的概率,

## 发射概率

- $p(x_n | z_n, \phi)$

  $$p(x_n | z_n, \phi) = \prod_{k=1}^{K} p(x_n | \phi_k)^{z_{nk}}.$$

观察变量与潜在变量联合分布

- $$p(X, Z | \theta) = p(z_1 | \pi) \left[ \prod_{n=2}^{N} p(z_n | z_{n-1}, A) \right] \prod_{m=1}^{N} p(x_m | z_m, \phi).$$

## 生成式观点考虑 HMM

1. 选择 $z_1$ ，由 $\pi_k$ 控制，
2. 计算 $x_1$
3. 计算 $z_2$ 利用 $p(z_2|z_1)$

⋮