**一阶马尔可夫**

- 假设，每个事件相关只与最近一次的观察相关

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

**二阶马尔可夫**

- 为了让数据更关联

$$p(x_1,\dots,x_N)=p(x_1)p(x_2|x_1)\prod_{n=3}^{N}p(x_n|x_{n-1},x_{n-2}).$$

**M阶**

- $p(x_n|x_{n-M},\dots,x_{n-1})$.

**参数量**

- 一阶 $K(K-1)$

  M阶 $K^M(K-1)$

  其中K为每个变量可能取值.

**参数量 V.S. 数据影响**

- 引入额外的潜在变量来使得丰富的一类模型能够从简单的成分中构建.

$$p(x_1,\dots,x_N,z_1,\dots,z_N)=p(z_1)\prod_{n=2}^{N}p(z_n|z_{n-1})\prod_{n=1}^{N}p(x_n|z_n)$$

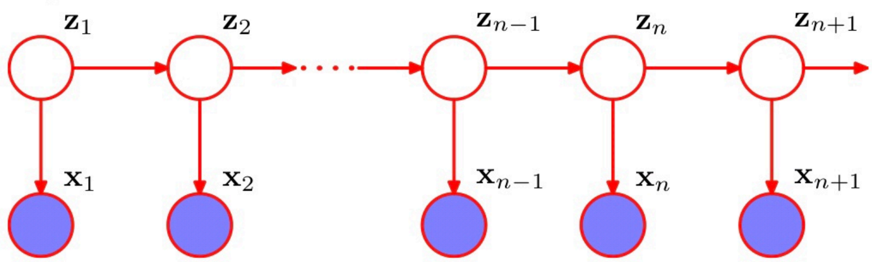
隐变量可行 why

使用d-划分准则，我们看到总存在一个路径通过潜在变量连接了任意两个观测变量$\boxed{x_n和x_m}$，并且这个路径永远不会被阻隔。因此对于观测变量$x_{n+1}$来说，给定所有之前的观测，条件概率分布$p(x_{n+1} \mid x_1, \dots, x_n)$不会表现出任何的条件独立性，因此我们对$x_{n+1}$的预测依赖于所有之前的观测。然而，观测变量不满足任何阶数的马尔科夫性质。我们在本章的后面几节会讨论如何计算预测分布。

① 任意 $x_n$ 和 $x_m$ 可达；

② $x_{n+1}$ 受到所有之前的影响.

隐变量




此图描述了两个重要模型：

① 潜在变量离散，则为HMM，

注意：HMM的观察变量可以离散或连续

② 若潜在变量和观察都为高斯，则为线性系统.

参数量。

$\theta = \langle \pi, A, \phi \rangle$

其中 $\pi$： 1

A：转移概率 $A_{jk} = p(z_{nk}=1 \mid z_{n-1,j}=1)$. $K(K-1)$.

$\phi$：发射概率 $p(x_n \mid z_n, \phi) = \prod_{k=1}^{K} p(x_n \mid \phi_k)^{z_{nk}}$.

具体讨论参见隐马尔可夫部分.