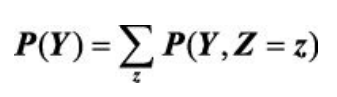
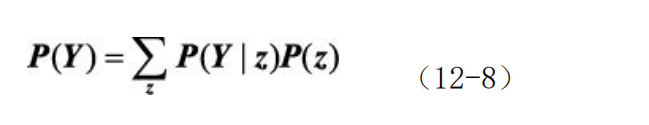
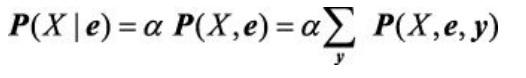
概率公理

边缘化/求和消元：

 条件化规则

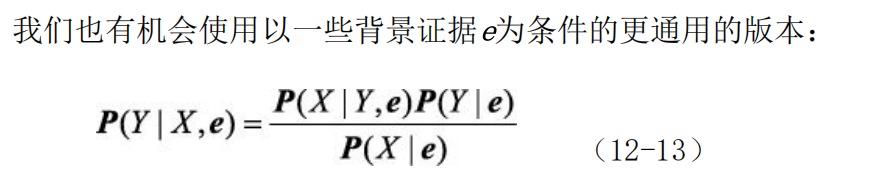
**归一化（使用来拟合分布进行推断的标准方法）**：



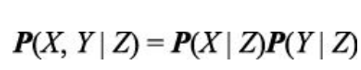
独立性：

如果整个**变量集能分解成独立子集**，则**完全联合分布可以分解成这些子集上的单独联合分布（的乘积）**

贝叶斯法则（以证据背景为条件）：



条件独立性：两个变量都与第三个变量有关，但给定第三个变量后，这两个变量毫无关联

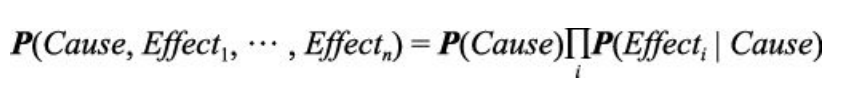


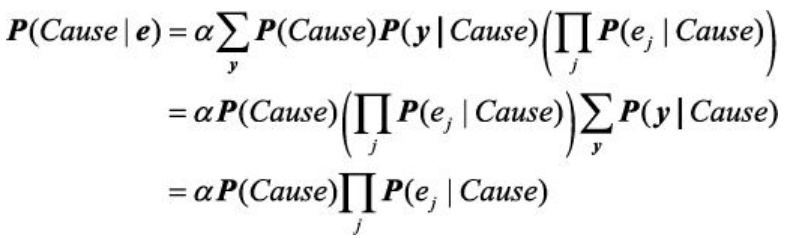
带归一化法则的贝叶斯法则：



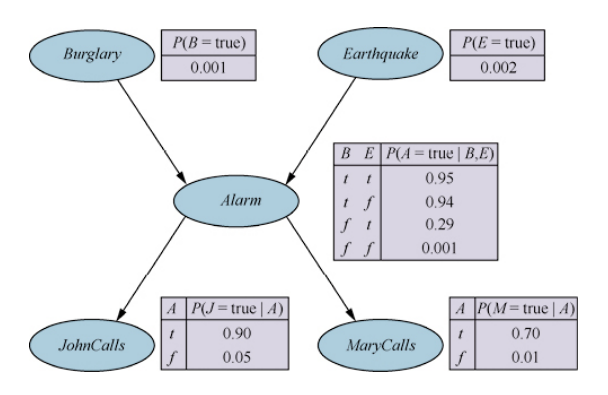
例，向量的对应位置相乘

朴素贝叶斯模型：假设所有结果（无论是已观测或未观测）在给定原因是时条件独立的





**贝叶斯网络**



每个结点表示一个随机变量

每结点都有一个CPT表示其在不同的父节点取值组合条件下的概率分布（一般只列出为真的条件概率）

**贝叶斯网络中的推断**

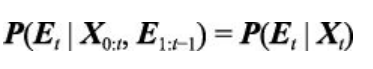
查询变量X，一组证据变量e，隐变量：在贝叶斯网络中但不是查询变量也不是证据的变量

**枚举推断：**查询P(B| j, m)：只用计算b的情况



（即贝叶斯网络的语义：父节点影响子节点，父父节点影响父节点）

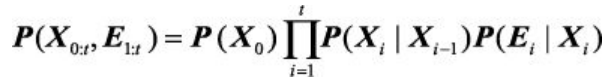
**时间上的概率推理**

马尔可夫转移模型：P(Xt| X0:t-1)=P(Xt| Xt-1) 后者即为转移模型

传感器模型： 后者即为传感器模型

（任何称职的状态都足以生成当前的传感器值）

初始状态+转移模型+传感器模型 = 任意时刻所有变量的联合分布



**滤波**：给定至今所有证据，计算当前状态的后验分布，即

前向消息**f1:k = P(Xk | e1:k)，(1:k)所有证据表明k时刻的状态情况**

滤波包含两个步骤， **初始化：f1:0 = P(X0)为对状态的先验信念**

1. 预测 **P(Xt+1 | e1:t) =**  **= ×f1:t （转移模型×前向消息）**
2. 更新 **f1:t+1 = P(Xt+1 | e1:t+1) = αP(et+1 | X t+1) P(Xt+1 | e1:t) = α传感器模型×f1:t**

**预测**：给定至今所有证据，计算未来状态的后验分布

（即不添加新证据的滤波）不停地乘以状态转移矩阵即可



**平滑**：给定至今所有证据，计算过去状态的后验分布

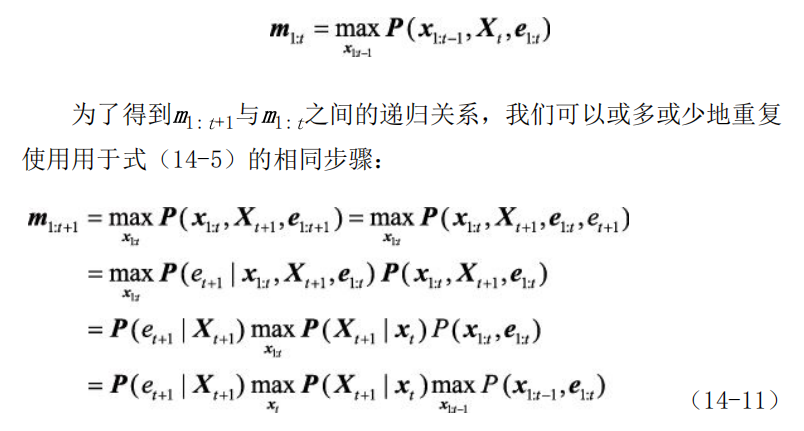
利用所有证据推断Xk = k之前的证据推断出Xk的概率 × Xk导致后续证据的概率

后向消息**bk+1:t = P(ek+1:t | Xk)，k时刻的状态情况对后续观测结果(k+1:t)的影响**

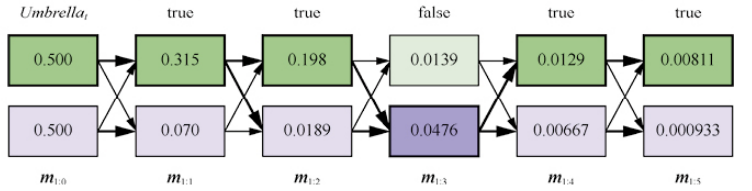
****初始化：**bt+1:t = P(et+1:t | Xt) = P( |Xt) = 1**

1. 计算反向消息：**P(ek+1:t | Xk) =**
2. ****再计算过去状态的后验分布：**P(Xk | e1:t)**

**最可能解释**：给定观测序列，找到最可能产生该序列的状态序列



从时刻0以先验 **m1:0 = P(X0)** 开始



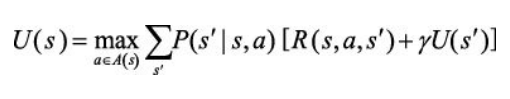
每次计算后选择概率大的状态，再计算下一状态。

**惯序决策**

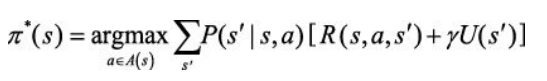
效用函数U(s)：从s出发的折扣效用的期望和

**贝尔曼方程：状态效用的计算公式**：A(s)为状态s时可采取的行动

状态的效用=该状态可采取的行动中：奖励+可能导致的下一状态效用 的期望 的最大值

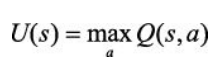


最优决策中动作的选择：arg表示使后式取max的动作a

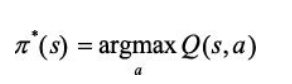


**动作效用函数/Q函数：**Q(s,a)在给定状态s下采取动作a的期望效用

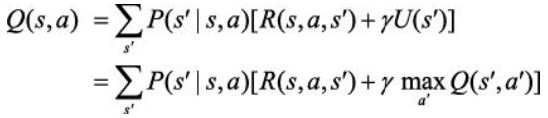
状态效用 = 该状态可采取的最优行动的效用值



最优策略的提取：



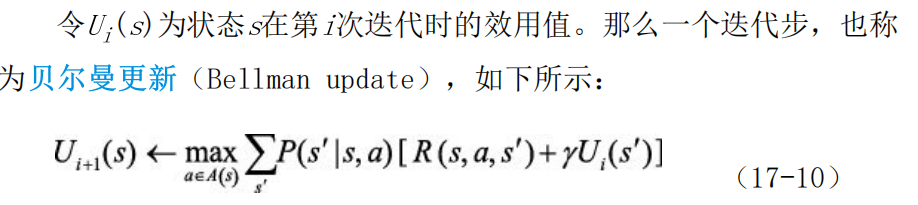
**Q函数的贝尔曼方程：动作效用的计算**



过求解关于U（或Q）的贝尔曼方程，我们可以得到所需的最优策略。（计算等式时最终使用的a即为π\*(s)）

**MDP求解算法**

**价值迭代**



每次价值迭代同时作用于所有状态，无限次更新后一定达到平衡，且一定是贝尔曼方程的（唯一）解，其对应的策略一定是最优的

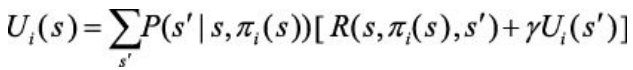
终止标准：**迭代过程在智能体的效用向量与其前一次的效用向量完全相同时终止。**

（双智能体零和博弈时只需一个智能体的效用向量不发生改变时即可终止）

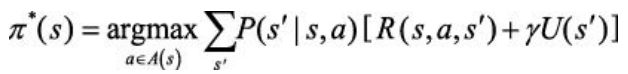
**策略迭代**

某个初始策略π0开始，交替执行：

1. 策略评估：对于策略πi，计算，即执行πi后每个状态的效用

（策略πi简化下的贝尔曼方程，省去了动作选择去最大值的步骤）

(2) 策略改进：使用基于Ui的一步前瞻，利用下式计算新的MEU策略πi+1



**判断题：**

如果奖励函数处处为零，那么所有的策略都是最优的