USTC

Chapter 13

Uncertainty

王子磊 (Zilei Wang)

Email: zlwang@ustc.edu.cn

http://vim.ustc.edu.cn

提纲

- 不确定性 (Uncertainty)
- 概率 (Probability)
- 语法和语义 (Syntax and Semantics)
- 推理 (Inference)
- 独立性 (Independence) 与 贝叶斯规则 (Bayes' Rule)

不确定性

行动 A_t = 起飞前 t 分钟前往机场 ➔ A_t 能否按时将我送达到机场?

问题：

1. 部分可观的（道路状态、其他司机的计划等等）
2. 有噪的传感器（交通报告）
3. 行动结果的不确定性（爆胎等）
4. 建模与交通流预测的极高复杂性

因而，一个纯粹的逻辑方法将遭受

1. 虚假风险：“A_{25} will get me there on time”，或者
2. 导出对决策来说太弱的结论

“A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc.”

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

处理不确定性的方法

- 默认 (default) 或 非单调 (nonmonotonic) 逻辑
  - 假设我的车不会爆胎
  - 假设 A_{25} 工作良好，除非与已知事实矛盾
  - 问题：哪些假设是合理的? 如何处理矛盾?
- Rules with fudge factors：
  - A_{25} $|\rightarrow_{0.3}$ get there on time
  - *Sprinkler* $|\rightarrow_{0.99}$ *WetGrass*
  - *WetGrass* $|\rightarrow_{0.7}$ *Rain*
  - 问题：组合问题的处理, e.g., *Sprinkler* causes *Rain*??
- 概率 (Probability)
  - 建模 agent 的置信度 (degree of belief)
  - 给定所有的已知事实 (available evidence)
  - “A_{25} will get me there on time with probability 0.04”

（模糊逻辑处理真值度 (degree of truth) 而不是不确定性，如 *wetGrass* 的真值度是 0.2）

概率

概率断言了综合效果 (summarize effects)

- 惰性 (laziness)： 难以列出所有规则和结论的完整集合等，且这种规则难以使用
- 无知 (ignorance)：缺少相关事实、初始条件等 (理论无知和实践无知)

主观 (Subjective) 或贝叶斯 (Bayesian) 概率：

- 概率将命题与 agent 自身知识状态相关联

  e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.06$

- 不对世界本身进行断言
- 命题概率随着新事实的出现会发生变化

  e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

不确定性下的决策

- 假设我们相信下面内容：

  $P(A_{25}$ gets me there on time | …) = 0.04

  $P(A_{90}$ gets me there on time | …) = 0.70

  $P(A_{120}$ gets me there on time | …) = 0.95

  $P(A_{1440}$ gets me there on time | …) = 0.9999

- 应该选取哪个行动?
  - 依赖于个人偏好 (preferences)： 错过航班 vs. 浪费时间
  - 效用理论 (Utility theory) 用于对偏好进行表达和推理

决策理论 (Decision theory) = 概率理论 (probability theory) + 效用理论 (utility theory)

语法 (Syntax)

- 基本要素：随机变量 (random variable)
- 与命题逻辑相似：可能世界是由对随机变量的赋值进行定义的
  - Boolean 随机变量

    e.g., *Cavity* (do I have a cavity?)
  - 离散随机变量

    e.g., *Weather* is one of <*sunny*,*rainy*,*cloudy*,*snow*>
- 样本空间的域值必须是完备的 (exhaustive) 、互斥的 (mutually)
- 基本命题通过单个随机变量的赋值进行构造
  - e.g., $Weather = sunny$, $Cavity = false$ (abbreviated as $\neg Cavity$)
- 复合命题由基本命题的逻辑连接构造
  - e.g., $Weather = sunny \vee Cavity = false$

语法 (Syntax)

- **原子事件**： 世界状态的一个完整规范 (complete specification)，Agent 对此可能是不确定的
  - E.g., 如果世界只有两个Boolean 变量 *Cavity* 和 *Toothache*， 那么有 4 个不同的原子事件：

$$Cavity = false \wedge Toothache = false$$

$$Cavity = false \wedge Toothache = true$$

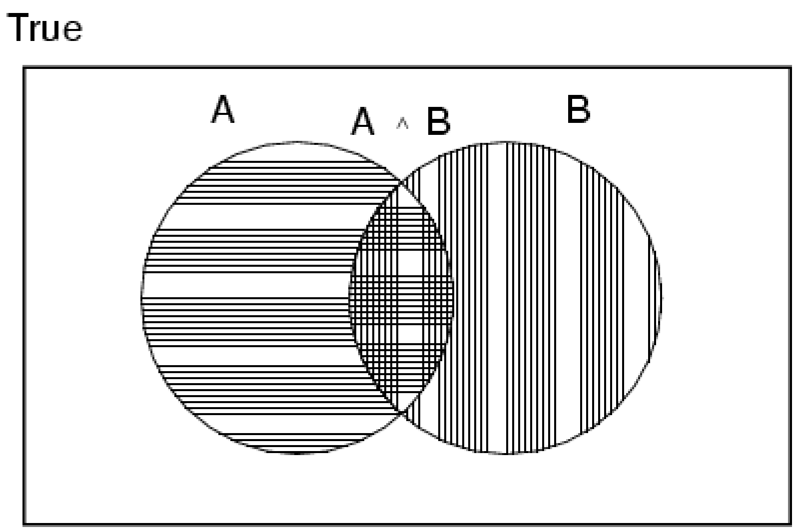
$$Cavity = true \wedge Toothache = false$$

$$Cavity = true \wedge Toothache = true$$

- 原子事件是互斥的和完备的

概率公理

- 对给定的两个命题 A, B
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



先验概率 (Prior probability)

- 命题的先验或无条件概率：

  e.g., $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$

  对应于获得任何证据之前的置信度

- 概率分布 (Probability distribution) 给出了所有可能赋值的概率

  $\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

- 联合概率分布 (Joint probability distribution) 对应于一组随机变量，给出了这些随机变量所构成的所有原子事件的概率

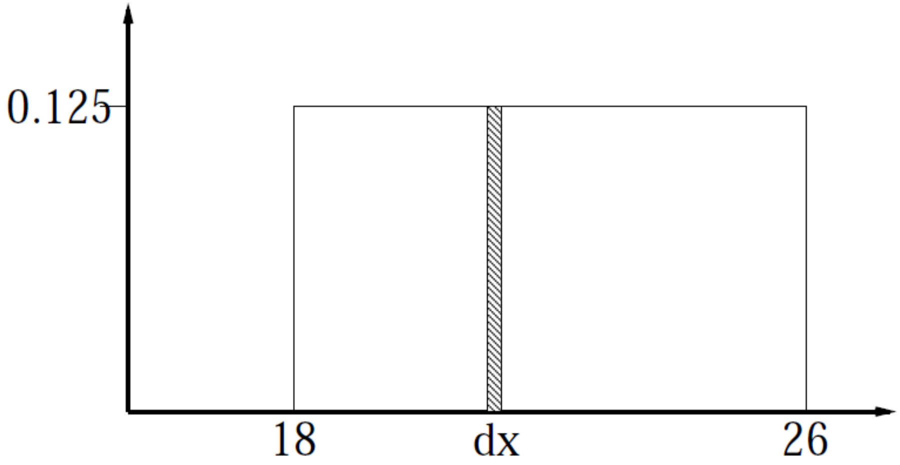
  $\mathbf{P}(Weather, Cavity) = a$ (4 × 2 matrix of values)

Weather =	sunny	rainy	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

- 每个关于该论域的问题都能够通过联合概率分布进行回答

连续变量的概率

- 将概率分布表示为一个参数化函数
  - $P(X = x) = U[18, 26](x)$ = uniform density between 18 and 26



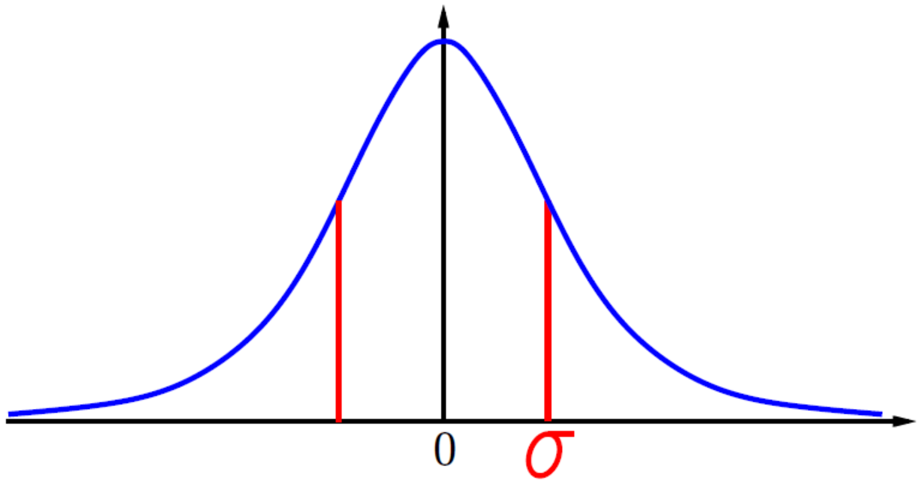
- 这里，P 是概率密度，积分为 1

$P(X = 20.5) = 0.125$ really means

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

高斯 (Gaussian) 分布

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$



条件概率 (Conditional probability)

- 条件或后验概率

  e.g., $P(cavity \mid toothache) = 0.8$

  i.e., given that *toothache* is all I know

- 关于条件概率的说明：

  $\boldsymbol{P}(Cavity \mid Toothache)$ = 2-element vector of 2-element vectors

- 如果我们知道更多，e.g., *cavity* is also given，那么我们有：

  $P(cavity \mid toothache, cavity) = 1$

- 新证据可能是无关的，通常进行简化，如：

  $P(cavity \mid toothache, sunny) = P(cavity \mid toothache) = 0.8$

- 在领域知识范围内，这类推理是至关重要的

条件概率 (Conditional probability)

- 条件概率的定义：

  $P(a \mid b) = P(a \wedge b) / P(b)$ if $P(b) > 0$

- 乘法规则 (Product rule) 提供了等价的一种形式

  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

- 一个在整体分布上的总体版本也是成立的, e.g.,

  $\mathbf{P}(Weather,Cavity) = \mathbf{P}(Weather \mid Cavity)\, \mathbf{P}(Cavity)$

  (View as a set of 4 × 2 equations, not matrix mult.)

- 链式法则 (Chain rule)： 连续应用乘法规则

$$\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}$$

枚举推理

❖ 从联合概率分布开始

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

❖ 对任一命题 φ，求和为真的原子事件：

- $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$
- 能够计算出该命题的概率

枚举推理

- 从联合概率分布开始

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- 对任一命题 φ，求和它为真的原子事件：
  - $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

枚举推理

- 从联合概率分布开始

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- 对任一命题 φ，求和它为真的原子事件：
  - $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

枚举推理

❖ 从联合概率分布开始

	$toothache$		$\neg\ toothache$	
	$catch$	$\neg\ catch$	$catch$	$\neg\ catch$
$cavity$	.108	.012	.072	.008
$\neg\ cavity$	.016	.064	.144	.576

❖ 也能够计算条件概率：

$$
\begin{aligned}
P(\neg cavity|toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\
&= \frac{0.016+0.064}{0.108+0.012+0.016+0.064} = 0.4
\end{aligned}
$$

归一化 (Normalization)

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

- 分母可以看做为归一化常数 (normalization constant) α

$$\mathbf{P}(cavity \mid toothache) = \alpha\, \mathbf{P}(cavity, toothache)$$

$$= \alpha\, [\mathbf{P}(cavity, toothache, catch) + \mathbf{P}(cavity, toothache, \neg catch)]$$

$$= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

$$= \alpha\, \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

通用思想：计算查询变量的概率分布，可以固定证据变量 (evidence variables)，然后在隐变量 (hidden variables) 上求和并归一化

枚举推理

X 表示所有随机变量，典型地，我们感兴趣的问题是

给定证据随机变量 E 的数值，

计算查询随机变量 Y 的后验联合概率分布

设隐变量为 $H = X - Y - E$，通过在隐变量上求和能够获得联合实体的总概率

$$\boldsymbol{P}(Y \mid E = e) = \alpha\, \boldsymbol{P}(Y, E = e) = \alpha\, \Sigma_h\, \boldsymbol{P}(Y, E= e, H = h)$$

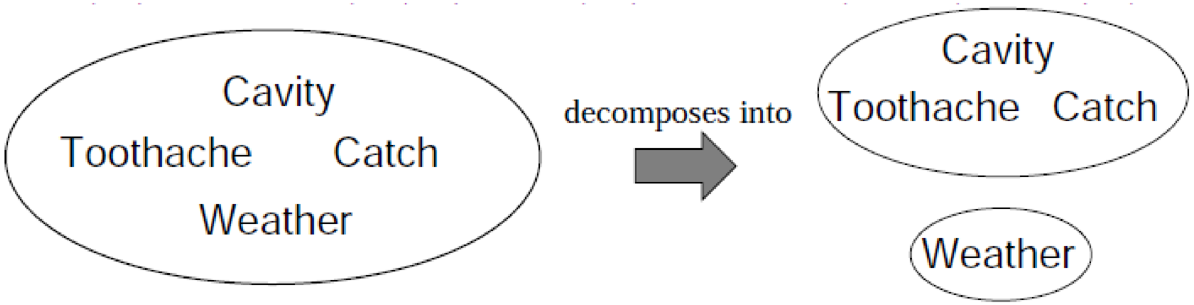
- 求和中的项是联合实体，因为 Y, E, H 在一起构成了完备空间

- **明显的问题：**
  1. 最坏情况的时间复杂性为 $O(d^n)$，这里 d 是最大的元数 (arity)
  2. 空间复杂度为 $O(d^n)$，用于存储联合概率
  3. 如何有效地找到 $O(d^n)$ 实体的数值?

独立性 (Independence)

- A 和 B 是独立的，当且仅当

  $P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)\, P(B)$



$\mathbf{P}(Toothache, Catch, Cavity, Weather)$

$= \mathbf{P}(Toothache, Catch, Cavity)\ \mathbf{P}(Weather)$

- 32 (8x4) 个实体降低为12(8+4)个；对 n 个独立的偏向硬币: $O(2^n) \rightarrow O(n)$
- 绝对独立性是强大的，但现实应用中很少
  - 牙医业 (Dentistry) 是一个具有成百个变量的大领域，它们并不独立，如何处理?

条件独立性 (Conditional independence)

- $\mathbf{P}(Toothache, Cavity, Catch)$ 有 $2^3 - 1 = 7$ 个独立实体

- 如果已知有 *cavity*，则探测到 *catches* 的概率与是否有 *toothache* 无关：

  (1) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$

- 如果已知没有 *cavity*，这种独立性依然存在

  (2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$

- 给定*Cavity*，*Catch* 相对于 *Toothache* 是条件独立的

  $\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$

- 一些等价的描述：

  $\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$

  $\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)$

条件独立性 (Conditional independence)

- 用链式法则写出完整的联合概率分布

  $\mathbf{P}(Toothache, Catch, Cavity)$

  $= \mathbf{P}(Toothache \mid Catch, Cavity)\, \mathbf{P}(Catch, Cavity)$

  $= \mathbf{P}(Toothache \mid Catch, Cavity)\, \mathbf{P}(Catch \mid Cavity)\, P(Cavity)$

  $= \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)\, \mathbf{P}(\mathrm{Cavity})$

  i.e., 2 + 2 + 1 = 5 个独立数字

- 在大部分情况下，使用条件独立性能够将联合概率分布的大小从 n 的指数级降到线性

- 条件独立性是不确定环境下最基本、健壮的知识形式

贝叶斯规则 (Bayes' Rule)

- 乘法规则 $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

  $\Rightarrow$ 贝叶斯规则 (Bayes‘ rule)： $P(a \mid b) = P(b \mid a)\, P(a) / P(b)$

- 或者是概率分布形式

  $$\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\, \mathbf{P}(Y) / \mathbf{P}(X) = \alpha\, \mathbf{P}(X|Y)\, \mathbf{P}(Y)$$

- 对根据因果概率评估诊断概率是非常有用的
  - $\mathbf{P}(Cause|Effect) = \mathbf{P}(Effect|Cause)\, \mathbf{P}(Cause) / \mathbf{P}(Effect)$
  - E.g., M 表示脑膜炎 (meningitis)，S 表示颈僵硬 (stiff neck)

    $P(m|s) = P(s|m)\, P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$

    （注意：脑膜炎的后验概率仍然是很小的）

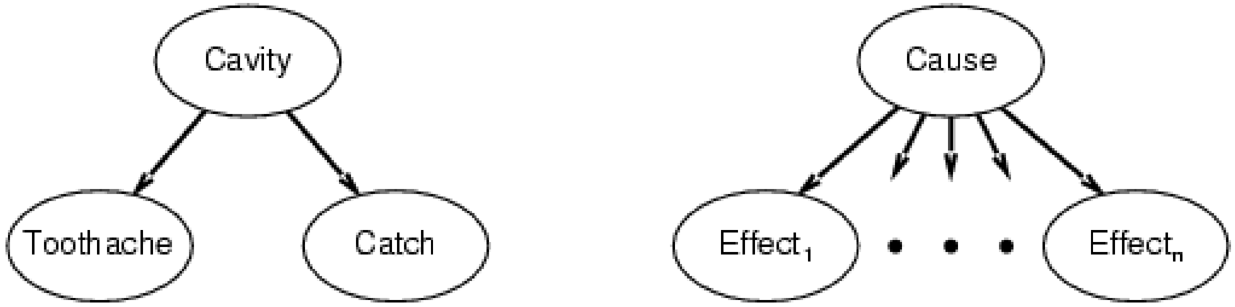
贝叶斯规则与条件独立性

$\mathbf{P}(Cavity \mid toothache \wedge catch)$

$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$

$= \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$

- 这是朴素贝叶斯 (naïve Bayes) 模型的一个典型实例
  - $\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$



- 参数量是 n 线性的

Wumpus World

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 **B** **OK**	2,2	3,2	4,2
1,1 **OK**	2,1 **B** **OK**	3,1	4,1

$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ iff $[i, j]$ is breezy

- 假定在概率模型中，我们只考虑 $B_{1,1}, B_{1,2}, B_{2,1}$

指定概率模型

- 完整的联合概率分布为 $\mathbf{P}(P_{1,1},\ldots,P_{4,4},B_{1,1},B_{1,2},B_{2,1})$
- 应用乘法规则 $P(Effect|Cause)$

$$\mathbf{P}(B_{1,1},B_{1,2},B_{2,1}\,|\,P_{1,1},\ldots,P_{4,4})\mathbf{P}(P_{1,1},\ldots,P_{4,4})$$

  - 第一项：当 pits 与 breezes 相邻时为 1，否则为 0
  - 第二项：pits 是随机放置的，每个方格的概率是 0.2，对 n 个 pits

$$\mathbf{P}(P_{1,1},\ldots,P_{4,4}) = \prod_{i,j=1,1}^{4,4}\mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

观测与查询

- 我们知道以下事实：

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$
$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

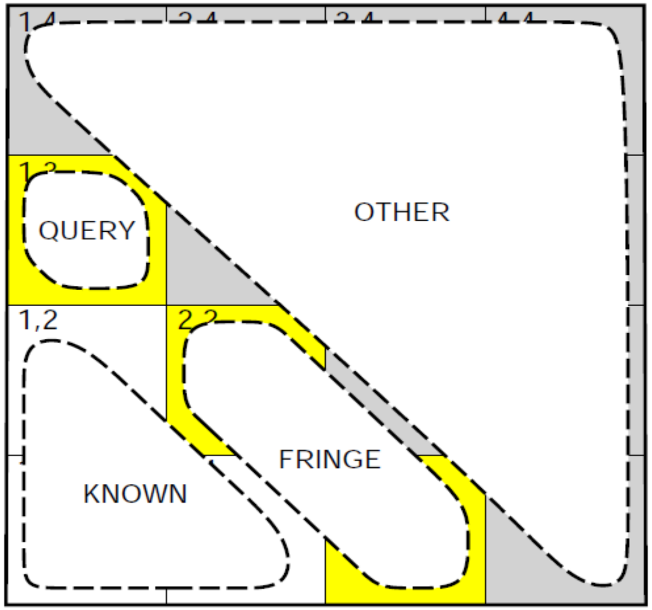
- 查询是 $\mathbf{P}(P_{1,3}|known, b)$
- 定义： $Unknown = P_{ij}$s other than $P_{1,3}$ and $Known$
- 采用枚举推理，则有：

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \Sigma_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

  - 与方格数呈指数级增长

使用条件独立性

- 基本观点：观测对除相邻隐方格之外的其他隐方格是条件独立的



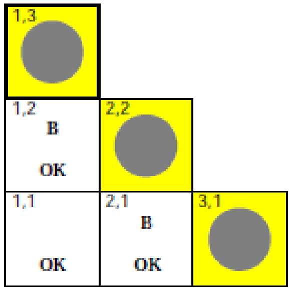
- 定义： $Unknown = Fringe \cup Other$

$$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$$

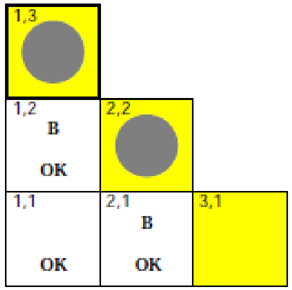
使用条件独立性

$$
\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) &= \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\
&= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
&= \alpha \, \underline{P(known)}\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \underline{\sum_{other} P(other)} \\
&= \alpha' \, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)
\end{aligned}
$$

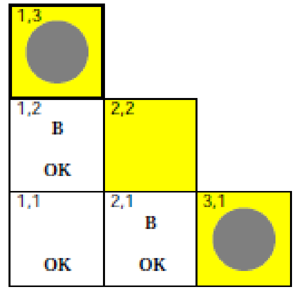
使用条件独立性



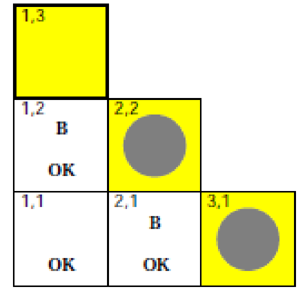
0.2 x 0.2 = 0.04



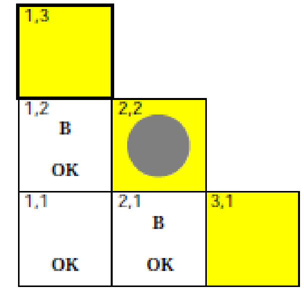
0.2 x 0.8 = 0.16



0.8 x 0.2 = 0.16



0.2 x 0.2 = 0.04



0.2 x 0.8 = 0.16

$$\begin{aligned}\mathbf{P}(P_{1,3}|known, b) &= \alpha' \langle 0.2(0.04+0.16+0.16),\ 0.8(0.04+0.16)\rangle \\ &\approx \langle 0.31, 0.69\rangle\end{aligned}$$

$$\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14\rangle$$

总结

- 概率是描述不确定知识的一种严格形式
- 联合概率分布给出了每个原子事件的概率
- 通过在原子事件上求和能够回答查询问题
- 对复杂领域，需要找到一种方法来降低联合概率的数目
- 独立性和条件独立性提供了重要工具

谢谢聆听！