USTC

Chapter 16

Rational Decisions

王子磊 (Zilei Wang)

Email: zlwang@ustc.edu.cn

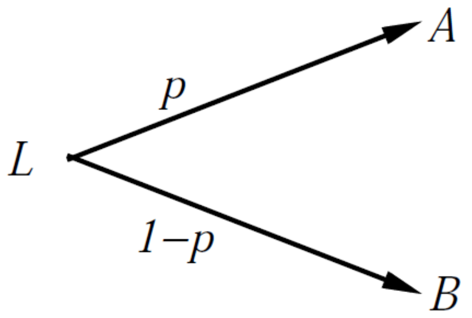
http://vim.ustc.edu.cn

提纲

- 理性偏好 (Rational preference)
- 效用函数 (Utilities)
- 金钱 (Money)
- 多属性效用函数 (Multiattribute utilities)
- 决策网络 (Decision networks)
- 信息价值 (Value of information)

偏好 (Preference)

- 一个Agent 要在 prizes (A, B, etc) 和 lotteries (即不确定的 prize) 之间进行选择



- Lottery $L = [p, A;\ (1-p), B]$

- 偏好记号：
  - $A \succ B$　　A preferred to B
  - $A \sim B$　　indifference between A and B
  - $A \succcurlyeq B$　　B not preferred to A

理性偏好

- 基本思想：理性 Agent 的偏好必须遵循某些约束
- 理性偏好：行动可描述为期望效用的最大化
- 约束：
  - 有序性 (Orderability)：$(A \succ B) \vee (B \succ A) \vee (A \sim B)$
  - 传递性 (Transitivity)：$(A \succ B) \wedge (B \succ C) \;\Rightarrow\; (A \succ C)$
  - 连续性 (Continuity)：$A \succ B \succ C \;\Rightarrow\; \exists p \;\; [p, A;\; 1-p, C] \sim B$
  - 可替换性 (Substitutability)：$A \sim B \;\Rightarrow\; [p, A;\; 1-p, C] \sim [p, B; 1-p, C]$
  - 单调性 (Monotonicity)：$A \succ B \;\Rightarrow\; (p \geq q \;\Leftrightarrow\; [p, A;\; 1-p, B] \succsim [q, A;\; 1-q, B])$
  - 可分解性 (Decomposability)：

$$[p, A; 1\text{-}p, [q, B; 1\text{-}q, C]] \sim [p, A; (1\text{-}p)q, B; (1\text{-}p)(1\text{-}q), C]$$

理性偏好

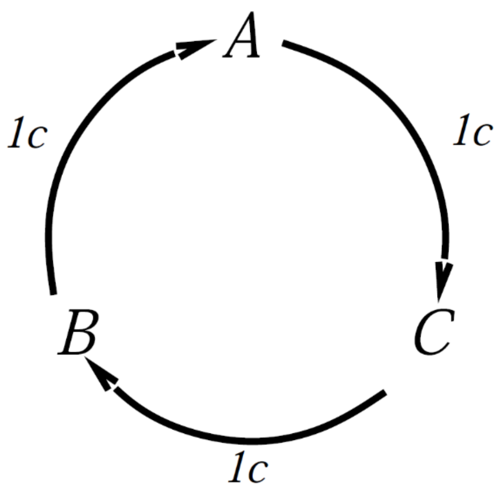
- 违反这些约束可能会导致明显的不合理性

- 例如：如果一个 Agent 不遵循偏好的传递性，它将总是损失的

If $B \succ C$, then an agent who has C would pay (say) 1 cent to get B

If $A \succ B$, then an agent who has B would pay (say) 1 cent to get A

If $C \succ A$, then an agent who has A would pay (say) 1 cent to get C



最大化期望效用 (MEU)

- 定理 (Ramsey, 1931; von Neumann and Morgenstern, 1944)：
  给定满足约束的偏好，存在一个实数函数 U (值函数)，满足：

$$U(A) \geq U(B) \quad \Leftrightarrow \quad A \succsim B$$
$$U([p_1, S_1;\ \ldots\ ;\ p_n, S_n]) = \textstyle\sum_i p_i U(S_i)$$

- MEU 原则：选择最大化期望效用的行动

注意：即使没有显式地表示或操作效用与概率，Agent 也可能是完全理性的（MEU）

- E.g., 游戏中的查找表

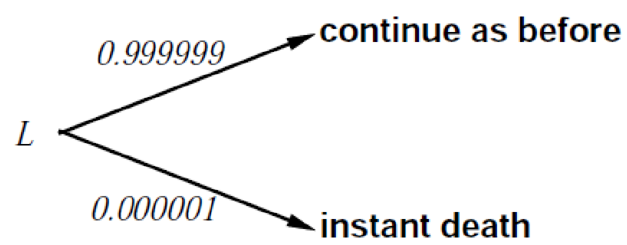
效用函数

- 效用是一个将状态映射为实数的函数
  - 什么样的实数？
- 人类效用评估的标准方法
  - 在一个给定的状态 A 和标准抽奖 L_p 之间做选择

    "best possible prize" $u_\top$ with probability p

    "worst possible catastrophe" $u_\perp$ with probability $(1-p)$
  - 调节抽奖概率 p，直到 $A \sim L_p$

pay $30 ~



效用尺度 (Utility scales)

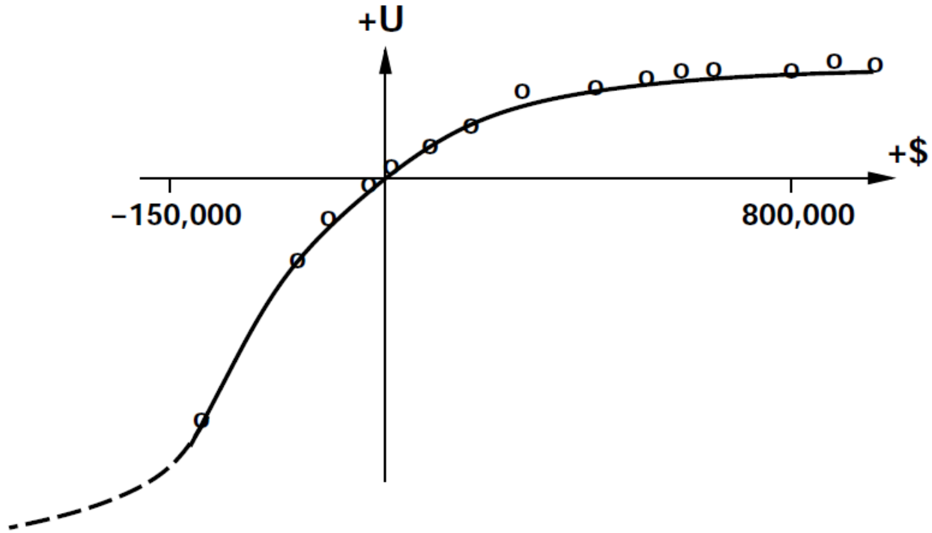
- 归一化效用：$u_\top = 1.0,\ u_\bot = 0.0$
- 微亡 (micromort)：百万分之一的可能性死亡
  - 俄罗斯轮盘赌，花大代价降低产品风险
- QALYs：quality-adjusted life years (质量调整寿命年)
  - 可能带来实质风险的医疗决策
- 注意：行为针对效用的线性变化具有不变性

$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$

➔ 在一个确定性的环境中，我们只需要顺序效用 (ordinal utility)，即只关心奖金的总顺序，具体数额是无关紧要的

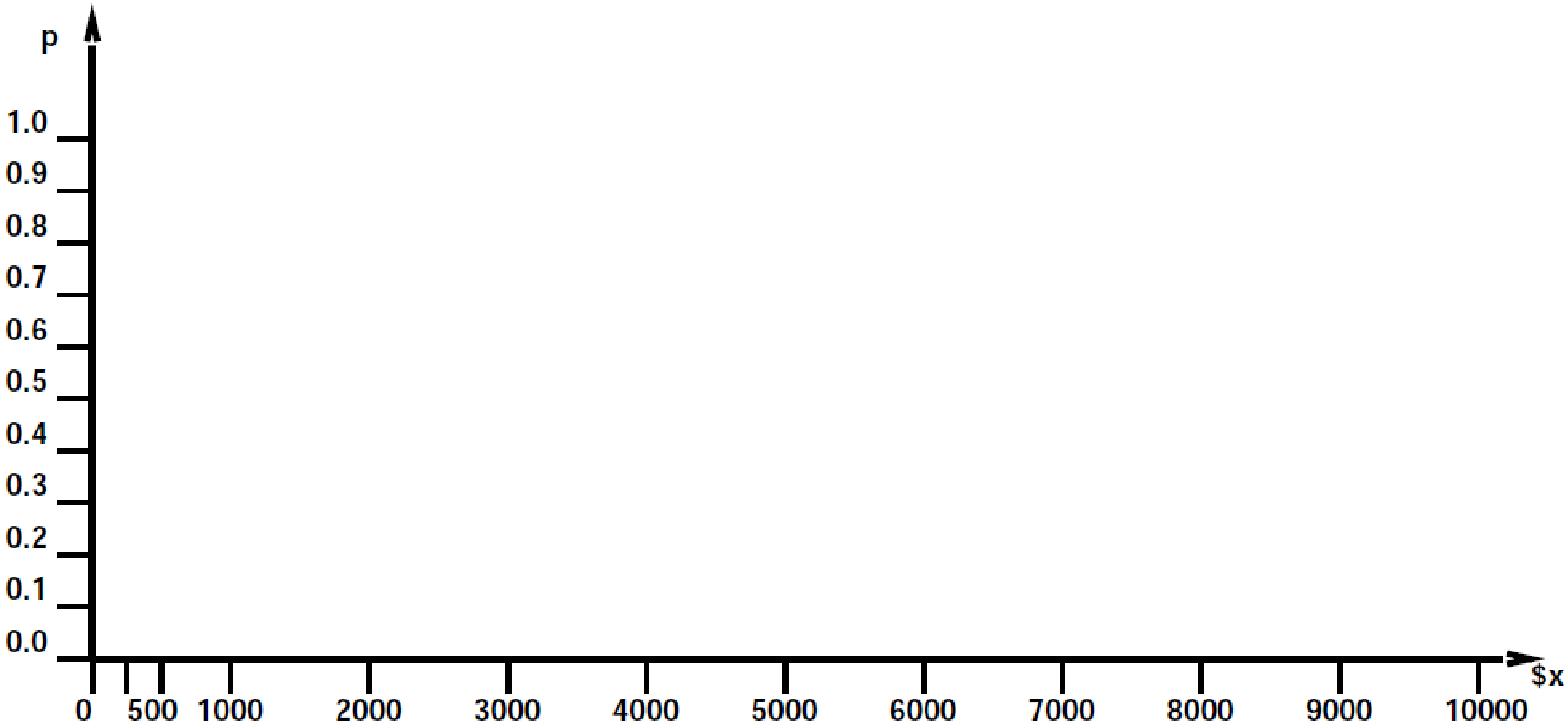
金钱 (Money)

- 金钱类似于效用，但在用法上与效用函数是 不同的
- 给定抽奖 L，具有期望货币价值 $EMV(L)$，通常 $U(L) < U(EMV(L))$
- 效用曲线：在奖金 x 和抽奖 $[p, \$M; (1-p), \$0]$ 之间无偏好下，关于 x 和概率 p 的变化曲线 (M 为一个大的固定值)
  - 趋险行为的典型效用曲线（对数正比）：



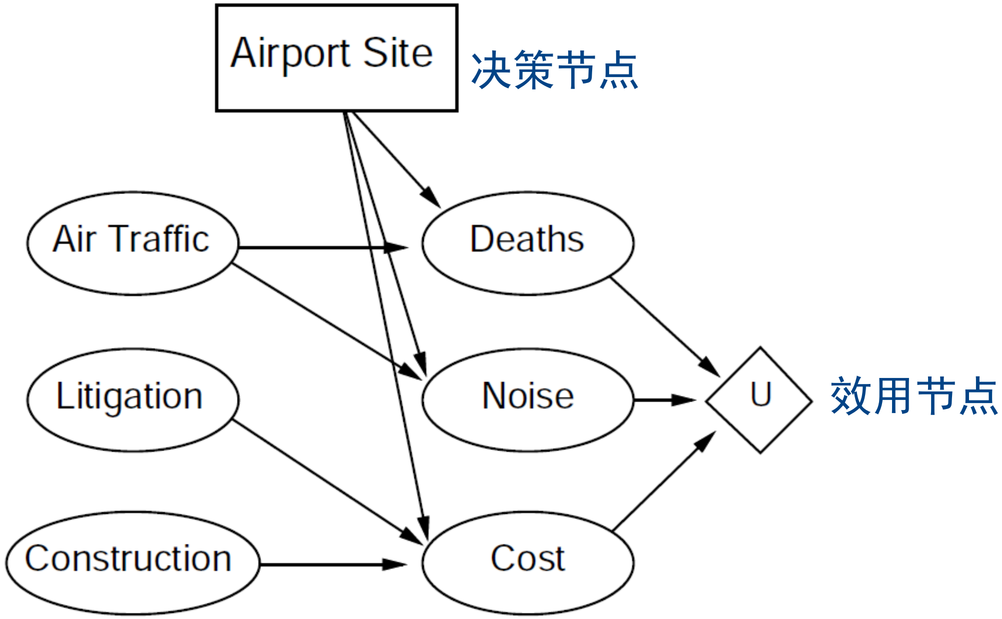
示例：Student group utility

- 对每个 x，调节 p 直到约半数同意进行抽奖 (M = 10000)



决策网络 (Decision networks)

❖ 添加决策节点和效用节点到置信网络中，使其能够进行理性决策



机场选址问题的一个简单决策网络

❖ 算法：针对每个行动节点的数值，计算给定行动和证据变量下的期望价值，最后选择 MEU 行动 （行动效用表， Q-function）

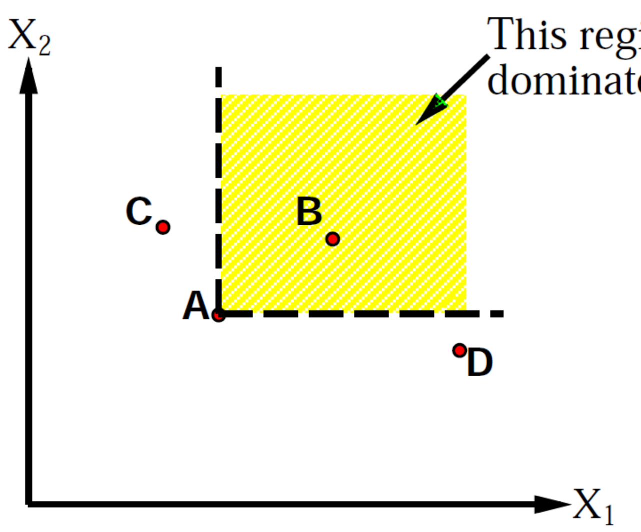
多属性效用函数

- 我们如何处理具有多变量 $X_1 \ldots X_n$ 的效用函数？
  - 如：$U(Deaths, Noise, Cost)$?
- 如何复合那些根据偏好行为评估的效用函数？
  - Idea1：找出那些无需 $U(x_1, \ldots, x_n)$ 完全表示的决策条件
  - Idea2：找到各种类型的偏好**独立性**，然后导出 $U(x_1, \ldots, x_n)$ 的规范化表示

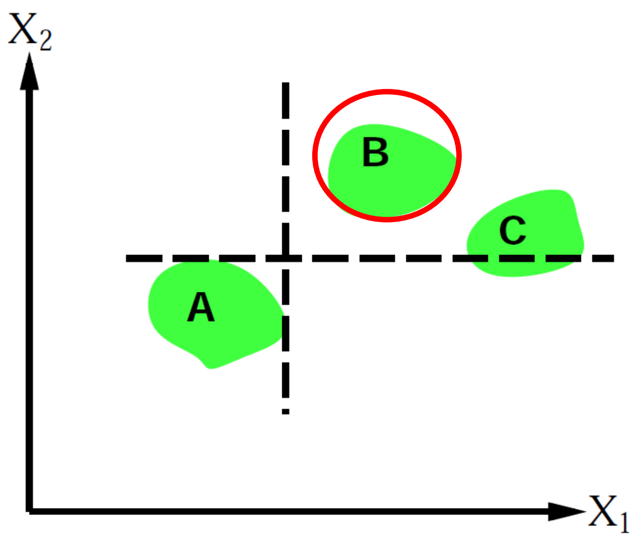
严格优势 (Strict dominance)

- 典型地，定义的 U 对每个属性都是单调的
- 严格优势：选择 B 严格由于选择 A，当且仅当

$$\forall i \quad X_i(B) \geq X_i(A) \qquad （因而 U(B) \geq U(A)）$$



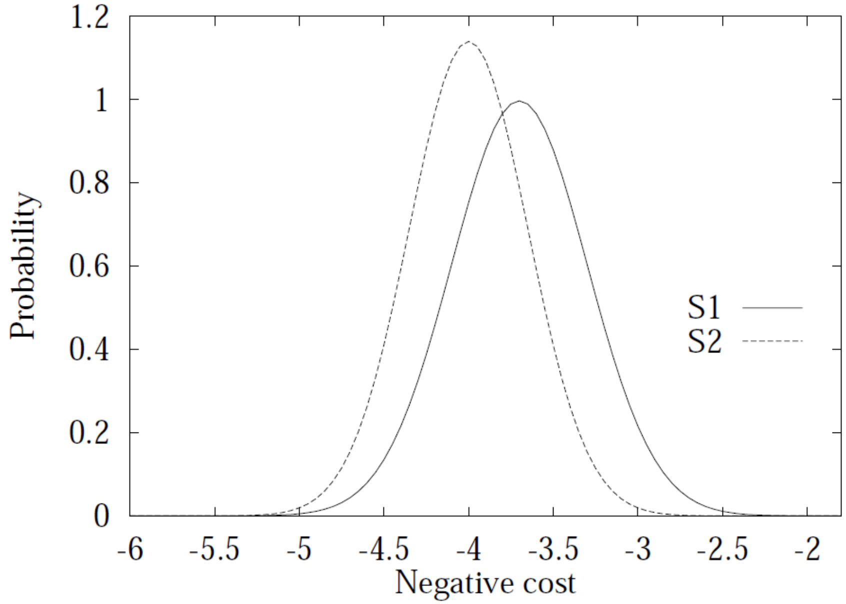
确定性属性

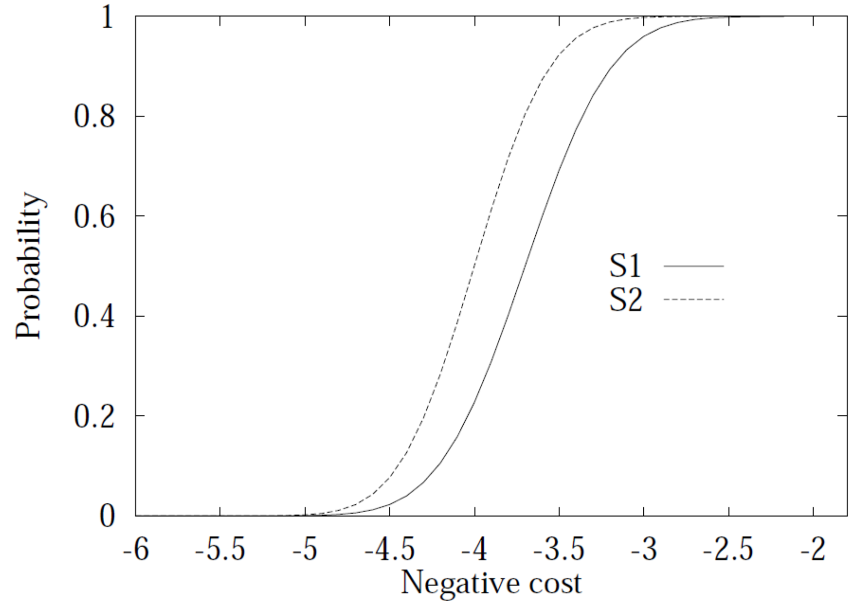


不确定属性

- 严格优势在实际应用中是很难满足的

随机优势 (Stochastic dominance)





- 分布 p_1 随机占优于分布 p_2，当且仅当 $\forall t \quad \int_{-\infty}^{t} p_1(x)dx \leq \int_{-\infty}^{t} p_2(t)dt$
- 如果 U 关于 x 是单调的，则结果分布为 p_1 的 A_1 随机占优于结果分布为 p_2 的 A_2：$\int_{-\infty}^{\infty} p_1(x)U(x)dx \geq \int_{-\infty}^{\infty} p_2(x)U(x)dx$
- 多属性情况：在每个属性上都是随机占优的 → 最优

随机优势 (Stochastic dominance)

- 随机优势常常能够通过定性推理进行判断，而无需分布的精确计算

- 例如：建造代价随着离城市距离的增大而增加

  S_1 离城市距离小于 S_2 ➔ S_1 在代价上随机占优于 S_2

- 例如：损失随着碰撞速度的提高而增加

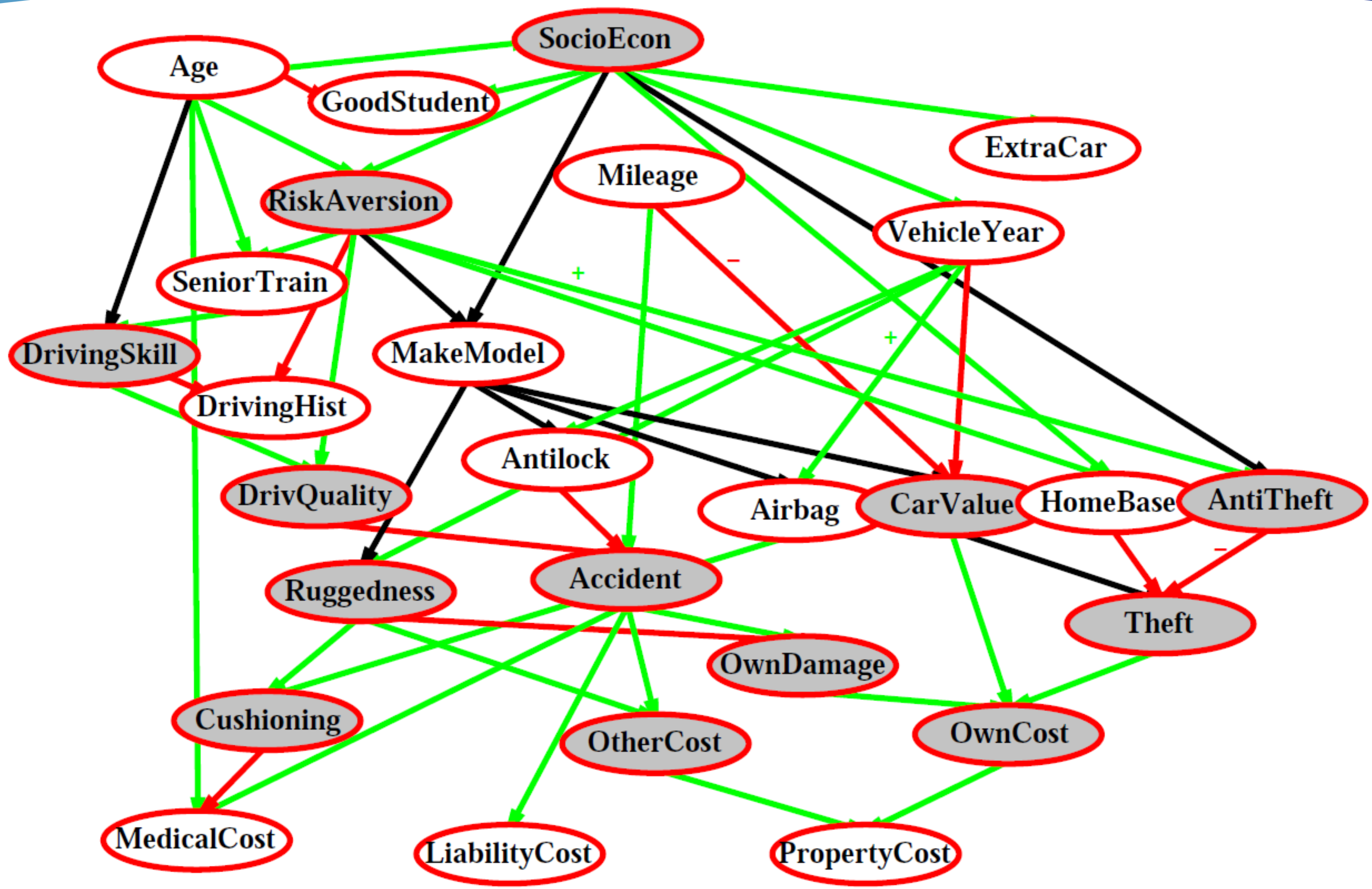
- 可以利用随机占优信息标注置信网络

  $X \xrightarrow{+} Y$ (X positively influences Y) 的意思是：

  "针对 Y 的其他父节点 $\mathbf{Z}$ 的每个值 $\mathbf{z}$,

  $\forall x_1, x_2 \;\; x_1 \geq x_2 \Rightarrow \mathbf{P}(Y|x_1, \mathbf{z})$ 随机占优于 $\mathbf{P}(Y|x_2, \mathbf{z})$ "

优势信息标注



SocioEcon
Age
GoodStudent
ExtraCar
Mileage
RiskAversion
VehicleYear
SeniorTrain
+
−
DrivingSkill
MakeModel
+
DrivingHist
Antilock
DrivQuality
Airbag
CarValue
HomeBase
AntiTheft
−
Ruggedness
Accident
Theft
OwnDamage
Cushioning
OtherCost
OwnCost
MedicalCost
LiabilityCost
PropertyCost

偏好结构：确定性的

- X_1 和 X_2 偏好独立于 (preferentially independent, PI) X_3，当且仅当 $\langle x_1, x_2, x_3 \rangle$ 和 $\langle x_1', x_2', x_3 \rangle$ 之间的偏好不依赖于 x_3
- E.g., $\langle Noise, Cost, Safety \rangle$

  ⟨20,000 suffer, \$4.6 billion, 0.06 deaths/mpm⟩ vs.
  ⟨70,000 suffer, \$4.2 billion, 0.06 deaths/mpm⟩

- 定理 (Leontief, 1947)：如果每对属性都偏好独立于它的补集，那么属性的每个子集都偏好独立于它的补集——相互偏好独立性
- 定理 (Debreu, 1960)：相互偏好独立性 ➔ 存在加性值函数

$$V(S) = \Sigma_i V_i(X_i(S))$$

因而，评估 n 个单属性的函数通常是比较好的近似

偏好结构：随机的

- 需要考虑抽奖偏好

  X 效用独立于(utility independent, UI) Y 当且仅当
  在 X 上的抽奖偏好不依赖于 y

- 相互效用独立性 (MUI)：每个子集与它的补集是效用独立的 ➔
  存在乘性效用函数：

$$
\begin{aligned}
U = {} & k_1U_1 + k_2U_2 + k_3U_3 \\
& + k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_3k_1U_3U_1 \\
& + k_1k_2k_3U_1U_2U_3
\end{aligned}
$$

信息价值

- 基本思想：计算获取每个可能证据片段的价值

  可以直接利用决策网络进行计算

- 示例：购买石油开采权
  - 两块区域 A 和 B，只有一块是有油的，具有价值 k
  - 每块的先验概率是 0.5，且是互斥的
  - 当前每块的价格是 $k/2$
  - 某咨询顾问提供了关于 A 的精确评估，合理的价格是多少？

- 解决方案：计算信息的期望价值

  = 给定信息下最佳行动的期望价值 - 没有信息下最佳行动的期望价值

[咨询报告可能是“oil in A”或 “no ail in A”，概率是 0.5]

$$
\begin{aligned}
= [&0.5 \times \text{ value of “buy A” given “oil in A”} \\
&+ 0.5 \times \text{ value of “buy B” given “no oil in A”}] \\
&- 0 \\
= (&0.5 \times k/2) + (0.5 \times k/2) - 0 = k/2
\end{aligned}
$$

一个通用公式

- 当前的证据 E，当前最佳行动 α，可能的行动结果 S_i，潜在新证据 E_j

$$EU(\alpha|E) = \max_a \sum_i U(S_i)\ P(S_i|E,a)$$

- 假设我们已知 $E_j = e_{jk}$，则选择 $\alpha_{e_{jk}}$

$$EU(\alpha_{e_{jk}}|E, E_j = e_{jk}) = \max_a \sum_i U(S_i)\ P(S_i|E, a, E_j = e_{jk})$$

E_j 是当前并不知道数值的一个随机变量

➔需要计算所有可能值上的期望收益

$$VPI_E(E_j) = \left(\sum_k P(E_j = e_{jk}|E) EU(\alpha_{e_{jk}}|E, E_j = e_{jk})\right) - EU(\alpha|E)$$

(VPI = value of perfect information)

VPI 的属性

- 非负性 (Nonnegative)：关于期望价值，而不是真实价值

$$\forall j, E \quad VPI_E(E_j) \geq 0$$

- 非加性 (Nonadditive)：如获取两次 E_j

$$VPI_E(E_j, E_k) \neq VPI_E(E_j) + VPI_E(E_k)$$

- 顺序独立性 (Order-independent)

$$VPI_E(E_j, E_k) = VPI_E(E_j) + VPI_{E,E_j}(E_k) = VPI_E(E_k) + VPI_{E,E_k}(E_j)$$

注意：当多于一个证据片段能够获得时，依次最大化 VPI 进行选择并不总是最优的

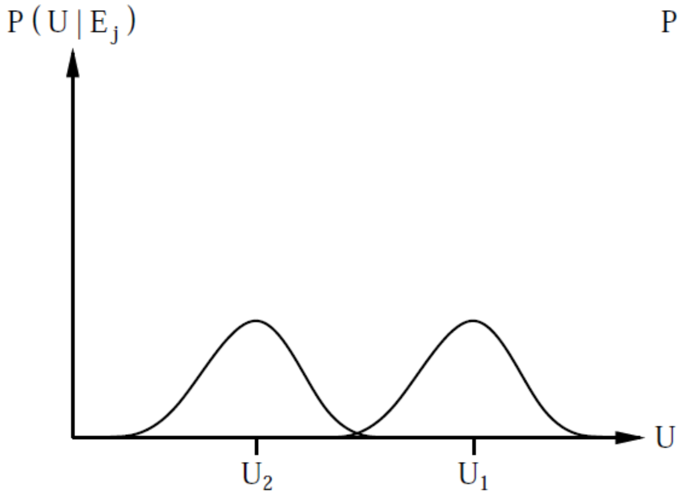
➔ 证据收集是一个序列决策问题

信息价值的定性分析

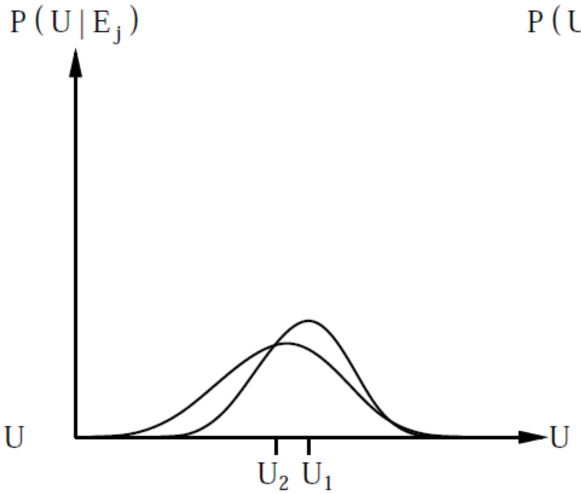
(a) 选择是明显的，信息几乎没有价值

(b) 选择是不明显的，信息很有价值

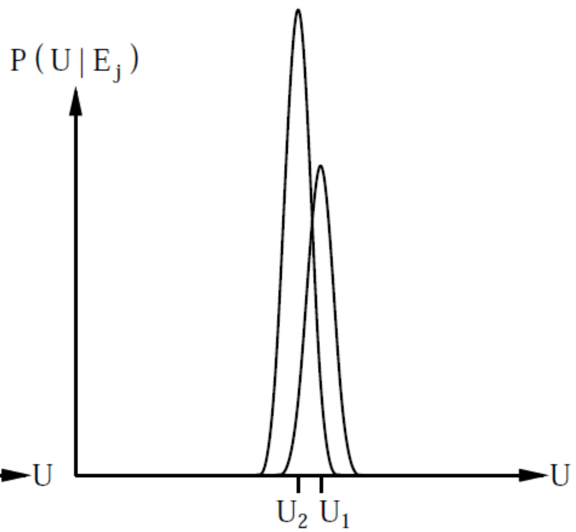
(c) 选择是不明显的，信息几乎没有价值



(a)



(b)



(c)

谢谢聆听！