制定简单决策
MAKING SIMPLE DECISIONS

张玲玲
计算机学院
zhanglling@xjtu.edu.cn

主要内容

1. 理性偏好
2. 效用函数
3. 金钱的效用
4. 多属性效用函数

制定决策

- 我们希望Agents能够根据他们的信念和意愿做出理性的决策。
- 例子：给机器人一个规划问题：我需要咖啡
  - 但是咖啡机坏了：机器人报告“No Plan！”
- 我们真正想要的是更鲁棒的行为：
  - 机器人知道如果我的主要目标无法实现该怎么办
  - 需要向机器人提供一些我对替代品偏好的指示

    例如，咖啡优于茶，茶优于水，水优于没有水等
- 但是这将更复杂：
  - 可以等45分钟修好咖啡机
  - 现在喝茶？等45分钟之后的咖啡？哪一种更好？
  - 可以表达对<饮料，时间>对的偏好

制定决策

- 制定决策：在替代方案中做出选择

- 这是困难的，原因是：
  - 决策复杂度
  - 存在不确定性
  - 在这些环境中存在多个目标，有时甚至是相反的目标

- 偏好（Preference）可以指导决策制定，对于在复杂情况下做出明智的选择至关重要。

行为的不确定性输出结果

- 行为（Action）可能产生不确定的输出结果（Outcome）。
  - 一次有200张彩票的彩票抽奖将向中奖者支付$1000。
  - 行为：花$10购买一张彩票
  - 结果：{win, not-win}
- 结果有一个概率分布: (0.005, 0.095)
- 每个结果都与某个效用（Utility）关联
  - Win：收获 $990
  - Not-win：损失 $10
- 是否应该执行这个行为?

理性决策

- 一个理性决策应该考虑：
  - 结果的相对重要性

    效用理论
  - 达到结果的置信度

    概率理论

- 决策理论 = 概率理论 + 效用理论
  - “当且仅当它选择产生最高期望效用的行为（在行为的所有可能结果中平均）时，Agent才是理性的。”

理性决策

- 一个不确定行为A有N种可能输出状态

$$Result_i(A), \qquad i=1,\cdots,N$$

- 每个状态被达到的概率是

$$P(Result_i(A)|Do(A),E) \quad (E\text{是部分观测})$$

例如：$P(win|buy, what\ we\ know\ about\ the\ lottery)$

- 一个事件（世界）状态 S 的相对重要性可由一个实值效用函数$U(S)$描述。
- 行为A的期望效用(Expected utility)定义为：

$$EU(A|E)=\sum_i P(Result_i(A)|Do(A),E)\,U(Result_i(A))$$

它是输出结果的加权平均效用值，输出结果的发生概率是权值。

最大期望效用原则（MEU）

- 理性Agent应该选择能够最大化期望效用的行为A：
$$Action = \underset{A}{\operatorname{argmax}} EU(A|E)$$
- 如果效用反映了性能指标，那么MEU将平均给出最高的性能得分。
- 某种意义上，可以认为MEU原则定义了人工智能的全部。一个Agent要做的所有事情就是计算各种量值，在其行动上最大化效用，然后采取行动。
- MEU原则的思想是“Agent应该做正确的事情”。

为什么使用效用？

- 为什么一个理性Agent要遵守MEU？为什么最大化如此特别的平均效用？

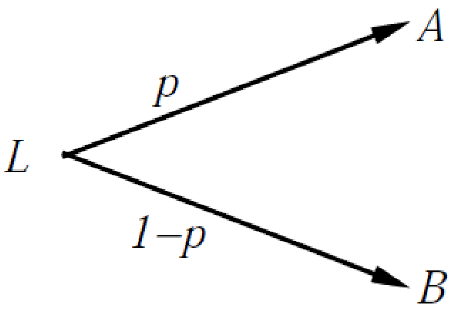
- 替代方案：
  - 最大化所有可能效用的加权立方和
  - 最小化最坏的可能损失
  - 最大化中位数效用
  - …

理性偏好

- 最大期望效用原则能够从理性偏好的约束中推导出来。
- 彩票抽奖（Lottery）：每个行为的一组输出结果。结果是奖品，结果是偶然决定的。
- 一张彩票（Ticket）：每个行为。
- 一次抽奖Lottery L具有结果$S_1, \cdots, S_n$，发生概率分别是$p_1, \cdots, p_n$:
$$L = [p_1, S_1; p_2, S_2; \cdots; p_n, S_n]$$
- 每个结果S_i既可以是一个状态，也可以是再一次抽奖。
- 抽奖$[1, A]$也可写作A。
- 制定决策：在抽奖中做出选择。

理性偏好

- Agent在奖品（A, B等）和抽奖做出选择，因为抽奖时奖品是不确定的。
- Lottery $L = [p, A; (1-p), B]$



- Agent在抽奖或状态间的偏好定义为：
  - $A \succ B$　Agent偏好A甚于B
  - $A \sim B$　Agent对A和B偏好相同
  - $A \succsim B$　Agent偏好A甚于B或者偏好相同

效用理论的公理

- 偏好和彩票有一些常识性约束。
- 这些约束代表（或试图代表）我们通常所说的“理性行为”的含义。

- 想法：理性Agent的偏好必须服从约束。
- 理性偏好 ⇒ 可描述为期望效用最大化的行为。

效用理论的公理

- 理性Agent的偏好必须要遵守的6个约束
  1. 有序性
  2. 传递性
  3. 连续性
  4. 可替换性
  5. 单调性
  6. 可分解性
- 这些约束被认为是效用理论的公理

理性偏好的约束

- 有序性（Orderability）
  - 给定任意两次抽奖，一个理性Agent必须偏好于其中一次抽奖，或者认为对两者的偏好是一样的。也就是说，Agent必须做出决策，不能回避决策。

$$(A \succ B) \vee (A \prec B) \vee (A \sim B)$$

  - $(A \succ B), (A \prec B), (A \sim B)$，其中必须有一个而且只有一个成立。

- 传递性（Transitivity）
  - 给定任意三次抽奖，如果一个理性Agent偏好A甚于B，而且偏好B甚于C，那么该Agent一定偏好A甚于C。

$$(A \succ B) \wedge (B \succ C) \Longrightarrow (A \succ C)$$

理性偏好的约束

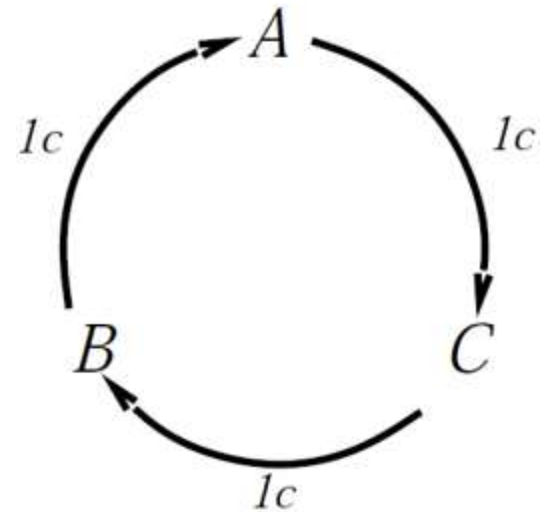
- 违反这些约束会导致不言而喻的不理性。
- 例如：可以通过使具有非传递性偏向的Agent交出其所有钱，从而说明传递性是必要的。
- 假设Agent具有非传递性偏向：$A \succ B \succ C \succ A$，其中$A$、$B$、$C$是可以自由交换的商品。

① 如果$B \succ C$，具有C的Agent愿意花一分钱交换B。

② 如果$A \succ B$，具有B的Agent愿意花一分钱交换A。

③ 如果$C \succ A$，具有A的Agent愿意花一分钱交换C。

④ 如此循环，直到Agent掏空所有钱。



不理性

理性偏好的约束

- 连续性（Continuity）
  - 如果某次抽奖B在偏好上介于A和C之间，那么一定存在某个概率p，使得该理性Agent在肯定得到B，与以概率p产生A并以概率$1\text{-}p$产生C的抽奖之间无偏向。

$$A \succ B \succ C \Rightarrow \exists p \quad [p, A; 1-p, C] \sim B$$

线性代数基本定理

- 单调性（Motononicility）
  - 假设两次抽奖有相同的两个可能结果A和B，如果一个Agent偏好A甚于B，那么该Agent一定偏好A的概率更高的抽奖，反之亦然。

$$A \succ B \Rightarrow (p \geq q \Leftrightarrow [p, A; 1-p, B] \succ [q, A; 1-q, B])$$

理性偏好的约束

- 可替换性（Substitutability）
  - 如果一个Agent在两次抽奖A和B之间无偏向性，那么该Agent在更复杂的两次抽奖之间也无偏向性。也就是说，这两次抽奖除了第一次中的A被B替换以外，其它都是一样的。这是成立的，而不用考虑抽奖中的概率和其它结果。

$$A \sim B \Rightarrow [p, A; 1-p, C] \sim [p, B; 1-p, C]$$

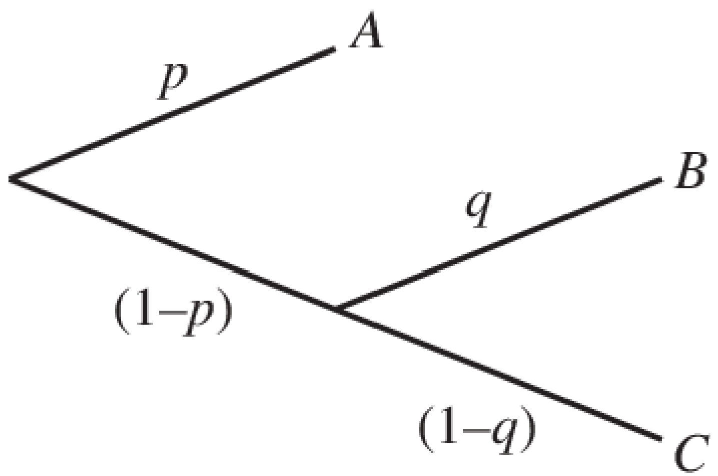
- 把其中的“$\sim$”替换成“$\succ$”后仍然是成立的：

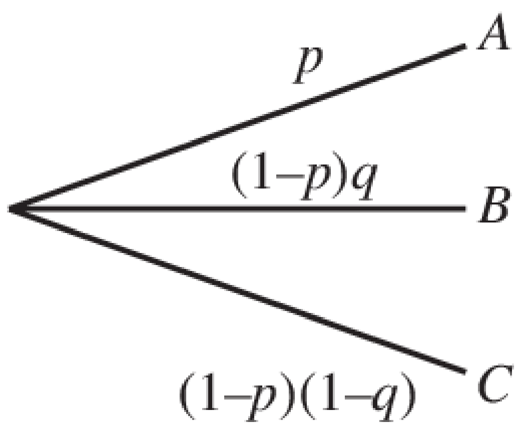
$$A \succ B \Rightarrow [p, A; 1-p, C] \succ [p, B; 1-p, C]$$

理性偏好的约束

- 可分解性（Decomposability）
  - 复合抽奖可以通过概率法则被简化为简单一些的抽奖。由于两次相继的抽奖能够被压缩成一个等价的单次抽奖，这曾被称为“赌博无乐趣”规则。

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





理性 vs. 非理性

- 任何违反上述公理的Agent的行为都是不理性的。

- 效用理论的公理实际上是关于偏好的公理——没有涉及效用函数！

主要内容

1. 理性偏好
2. 效用函数
3. 金钱的效用
4. 多属性效用函数

效用原则

- 效用函数的存在性定理：如果一个Agent的偏好遵守效用公理，那么存在一个实值函数U，满足：

$$U(A) > U(B) \Leftrightarrow A \succ B$$ （偏好A甚于B）

$$U(A) = U(B) \Leftrightarrow A \sim B$$ （在A和B之间无偏向）

- 抽奖的期望效用：一次抽奖的效用是把每个结果的概率乘以它的效用的乘积和：

$$U([p_1, S_1; \cdots; p_n, S_n]) = \sum_i p_i U(S_i)$$

$$EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))$$

最大化期望效用（MEU）

$$EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))$$

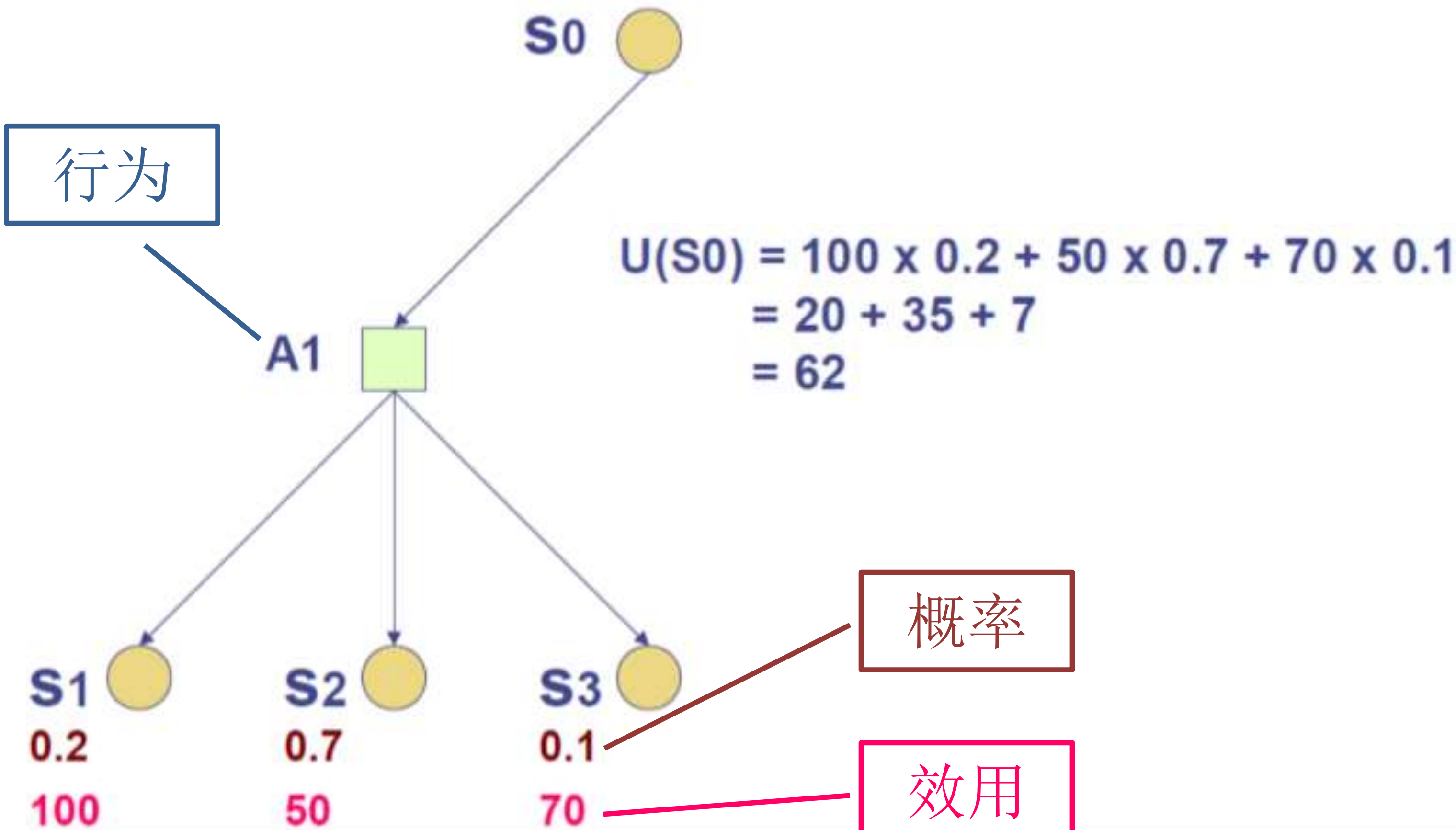
- 期望最大效用准则：理性Agent应该选择使期望效用最大的行为：

$$Action = \underset{A}{\operatorname{argmax}}\, EU(A|E)$$

- 如果效用反映了性能指标，那么MEU将平均给出最高的性能得分。

例子：行为的期望效用

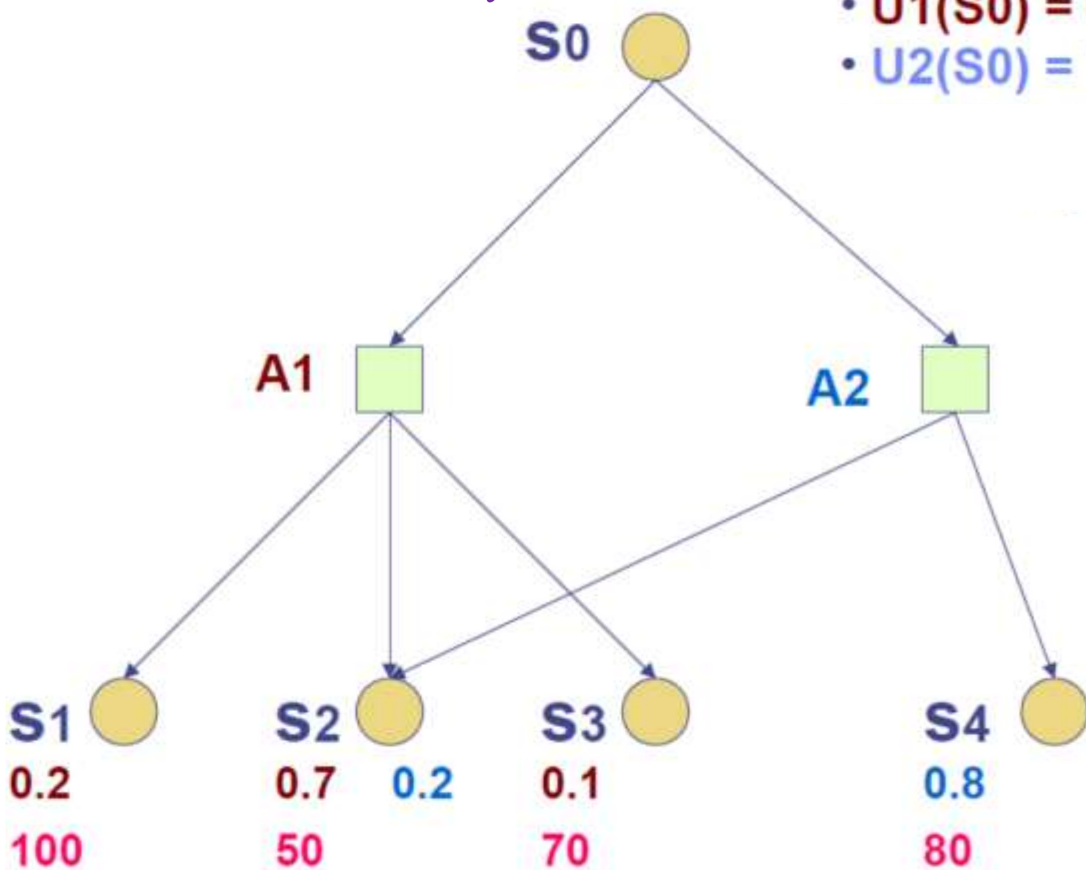
$$EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))$$



例子：行为的期望效用

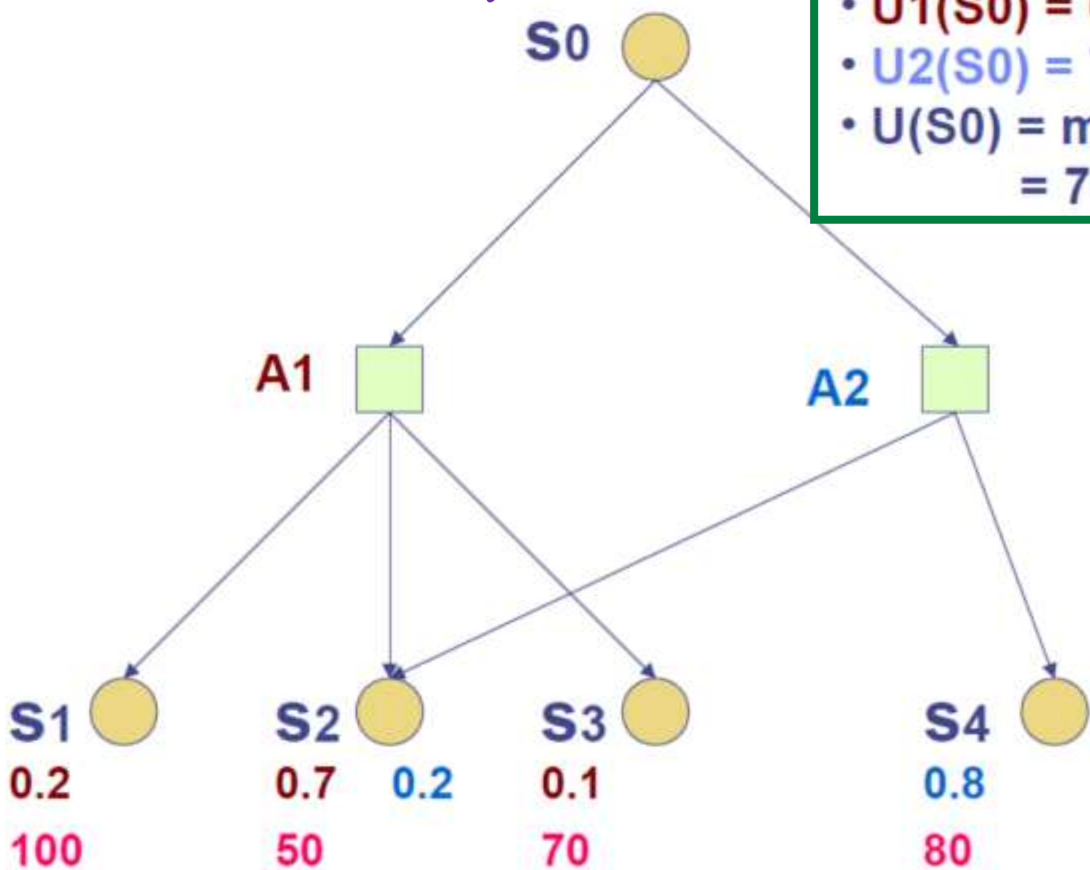
$$EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))$$

- U1(S0) = 62
- U2(S0) = 74



例子：行为的期望效用

$$EU(A|E) = \sum_i P(Result_i(A)|Do(A),E)\,U(Result_i(A))$$



- U1(S0) = 62
- U2(S0) = 74
- U(S0) = max{U1(S0),U2(S0)}
  = 74

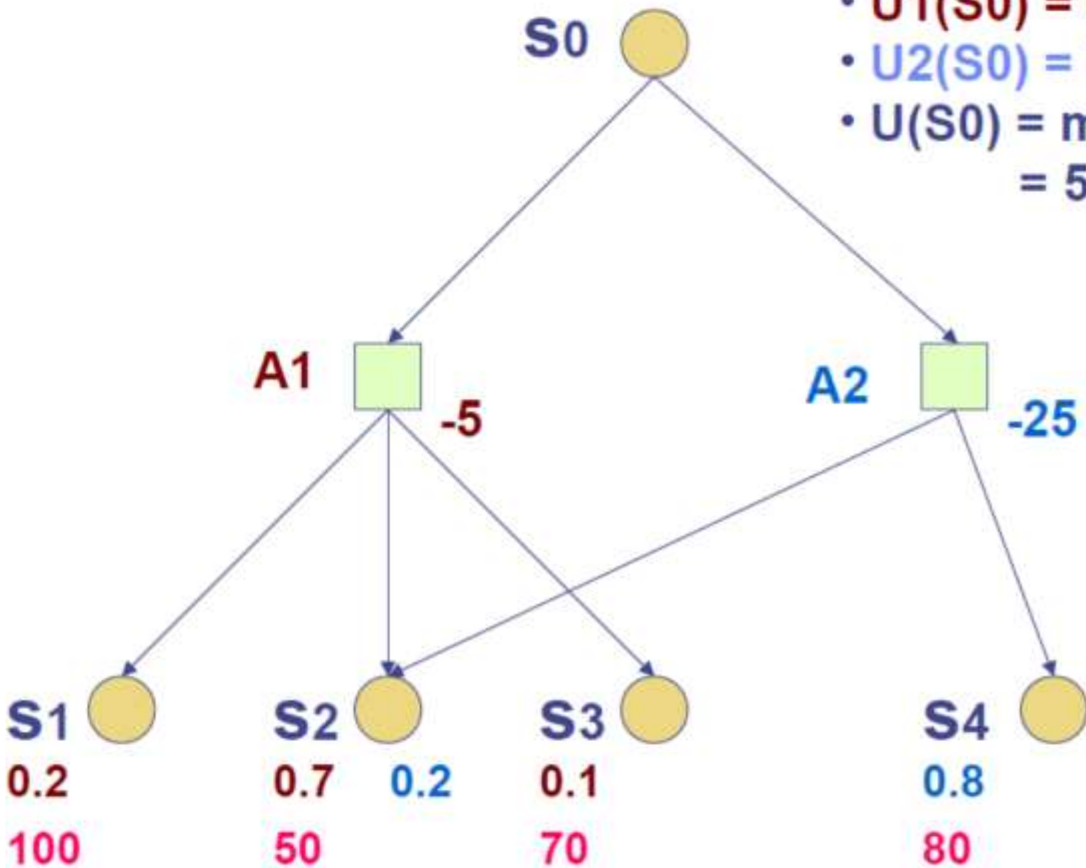
→ The agent should pick action 2.

例子：行为的期望效用

$$EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))$$

- U1(S0) = 62 – 5 = 57
- U2(S0) = 74 – 25 = 49
- U(S0) = max{U1(S0),U2(S0)}
  = 57

→ The agent should pick action 1.



效用函数是不唯一的

- 任何理性Agent都存在一个效用函数，但这个效用函数不是唯一的。
- 如果一个Agent的效用函数根据如下公式变换，它的行为将不会改变：
$$U'(S) = aU(S) + b$$
其中a和是b常数，且$a > 0$；这是一个仿射变换。
- 例子：效用类似温度， $U(S)$是摄氏温度， $U'(S)$是华氏温度：
$$U'(S) = 1.8 \times U(S) + 32$$

效用评估和效用尺度

- 效用是一个从抽奖映射到实数的函数。
- 效用函数是尺度无关的：
  - 在最大期望效用中，使用$U'(S) = aU(S) + b$与使用$U(S)$结果一致。
- 但是，建立某个尺度对记录和比较任何特定问题的效用是有帮助的。
- 通过固定两个特殊结果的效用可以建立一个尺度。例如：通过固定水的结冰点和沸点，建立温度的尺度。
- 归一化效用：
  - 效用“0”：“最坏的可能灾难”
  - 效用“1”：“最好的可能奖励”

效用评估和效用尺度

- 评估效用的一种方式是建立如下尺度：
  - $U(S) = u_{\top} = U(\text{"best possible prize"})$
  - $U(S) = u_{\perp} = U(\text{"worst possible catestrophe"})$
- 归一化效用：
  - $u_{\perp} = 0$， $u_{\top} = 1$
- 给定一个$u_{\perp}$和$u_{\top}$之间的效用尺度，任何特定奖励S的效用是通过让Agent在S和标准抽奖$L_p = [p, u_{\top}; (1-p), u_{\perp}]$之间选择进行评估的。
- 调整概率p直到Agent对S和L_p没有偏向性，即$S \sim L_p$ 。

主要内容

1. 理性偏好
2. 效用函数
3. 金钱的效用
4. 多属性效用函数

金钱的效用

- 黄牛花了三天时间通过网上预约购买了1双限量版球鞋。
  - A：张三现在愿意花￥3000购买球鞋。
  - B：黄牛通过微信朋友圈卖球鞋
    - 有80%概率卖到￥4000
    - 有20%概率卖不出去，即￥0
- 黄牛怎么办？

- 假设：效用 = 金额 = ￥
  - A：$EU(A) = U(S_{3000}) = 3000$
  - B：$EU(B) = p_{4000}U(S_{4000}) + p_0U(S_0) = 0.8 \times 4000 + 0.2 \times 0 = 3200$

金钱的效用

- 给你下面两张彩票之一：
  - A：20%概率中￥4000
  - B：25%概率中￥3000
- 你怎么办？

- 假设：效用 = 金额 = ￥
  - A：$EU(A) = p_{a1}U(S_{a1}) + p_{a2}U(S_{a2}) = 0.8 \times 0 + 0.2 \times 4000 = 800$
  - B：$EU(B) = p_{b1}U(S_{b1}) + p_{b2}U(S_{b2}) = 0.75 \times 0 + 0.25 \times 3000 = 750$

- 假设“金钱 = 效用”，是否合理？

金钱的效用

- 直接给你￥1,000,000 $EU(A) = 1{,}000{,}000$
- 或者让你投硬币 $EU(B) = 1{,}500{,}000$
  - 正面朝上：给你￥3,000,000
  - 背面朝上：给你0
- 你怎么选？
- 十万资产的人怎么选？
- 百万资产的人怎么选？
- 千万资产的人怎么选？
- 亿万资产的人怎么选？
- 马云怎么选？

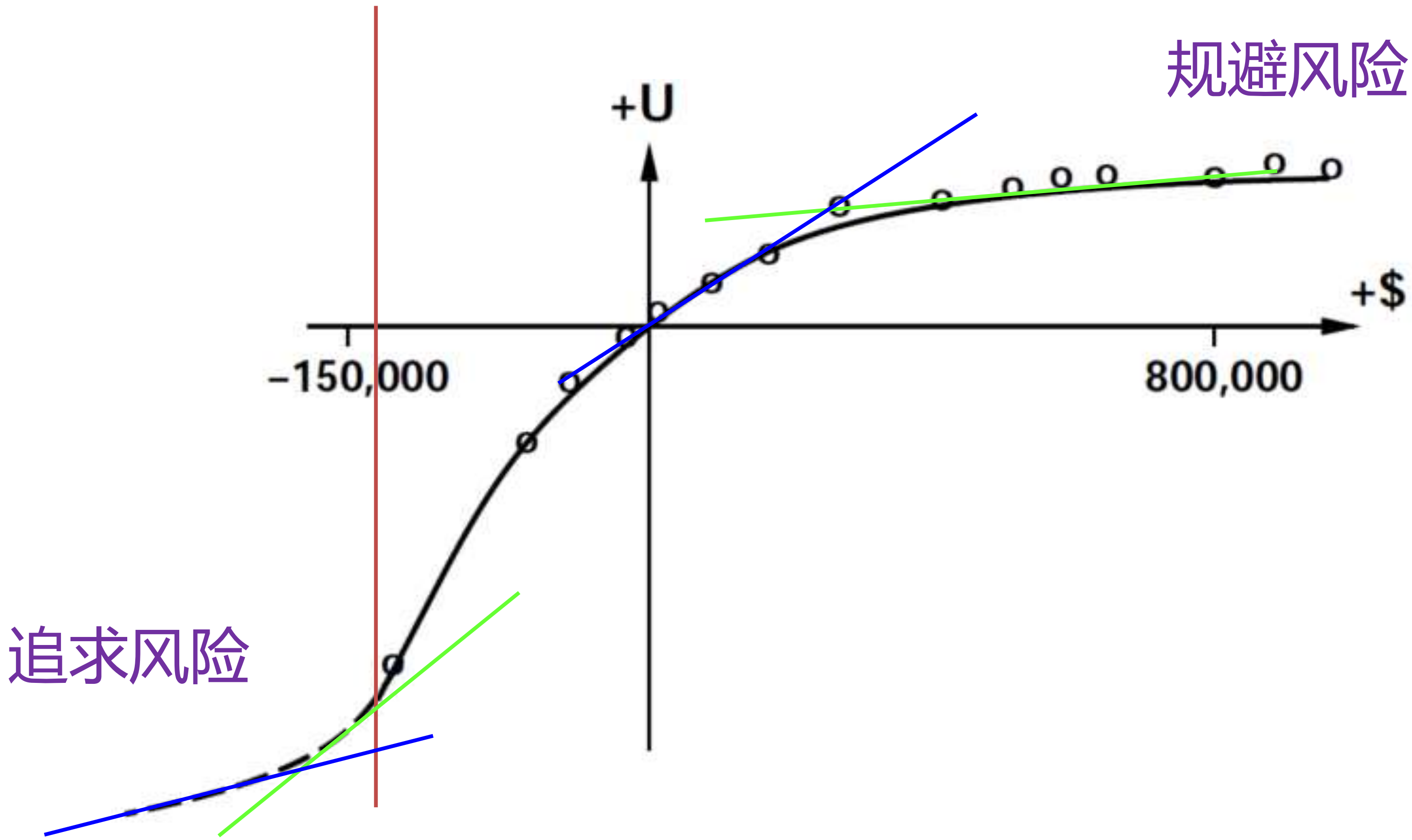
$p_{up} = 0.5$ $p_{down} = 0.5$

金钱的效用

- 人们认为金钱 (Money) 不是货币价值 (Monetary Value)！
- 创造效用理论的经济学家提出了以下建议：
  - 人们对金钱是单调性偏好（单调增）
  - 人们对金钱是非线性偏好（钱多？钱少？）
  - 金钱的效用与金钱的对数成正比！
  - 问题似乎是：追求风险与规避风险

金钱的效用

- 张三先生的金钱-效用曲线



金钱的效用

- 尽管货币价值有时候不能完全表征效用，有时还是使用货币价值（计算可用货币）作为效用函数。

- 期望货币价值（EMV, expected monetary value）是使用货币价值作为效用函数的期望效用：

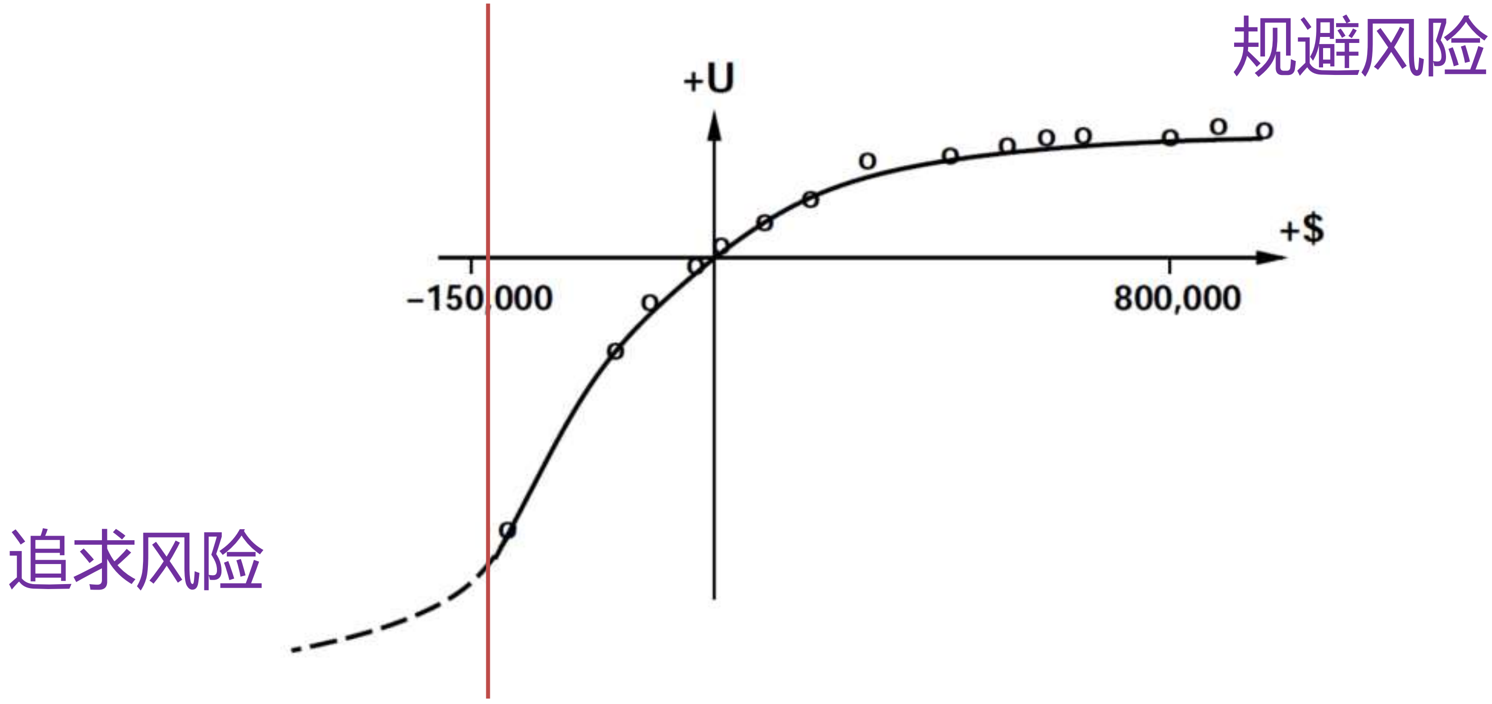
$$EMV(A) = \sum_i P\big(Result_i(A)\big) MV(Result_i(A))$$

$$\boxed{EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))}$$

金钱的效用

- **追求风险**：对于任意抽奖L，“面对这次抽奖”的效用大于“把这次抽奖的期望货币价值当作确定性的东西给你”的效用，即

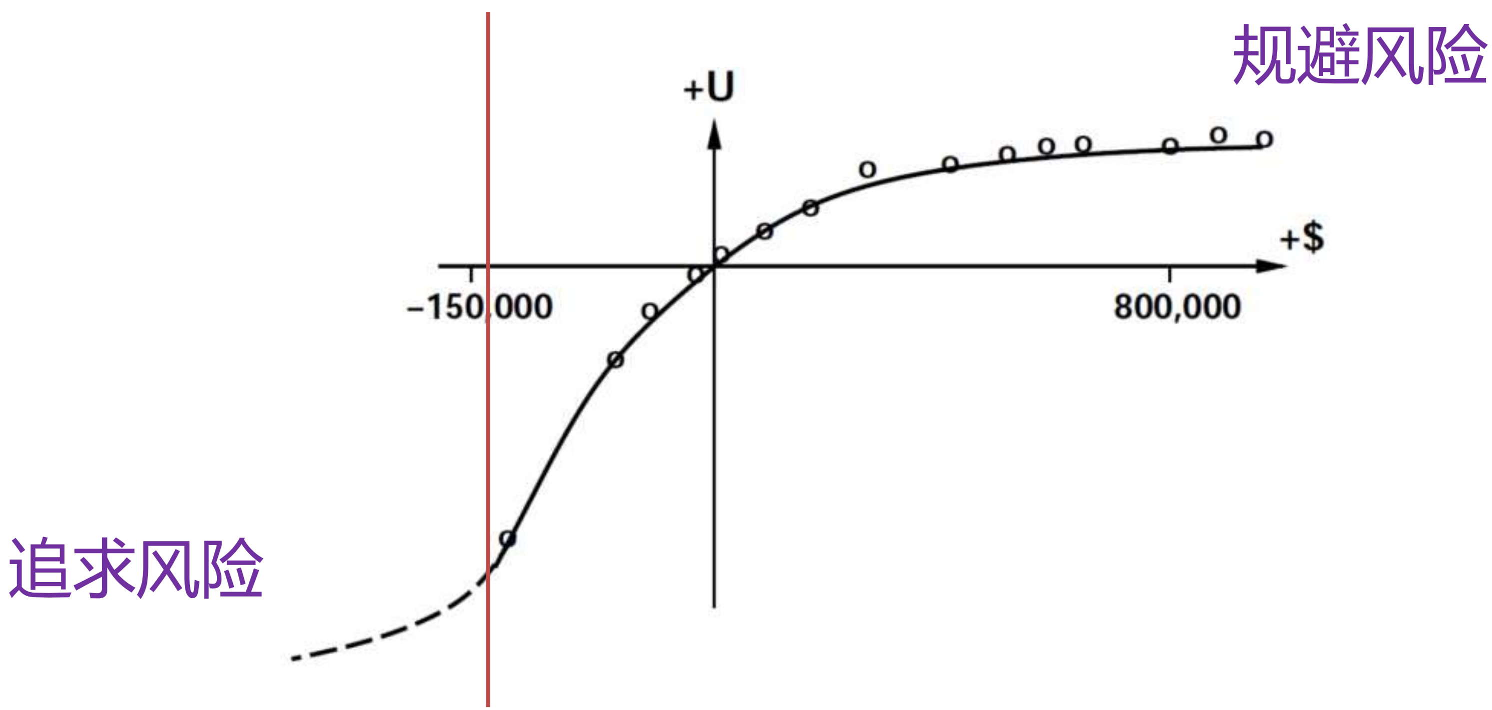
$$U(L) > U(S_{EMV(L)})$$



金钱的效用

- **规避风险**：对于任意抽奖L，“面对这次抽奖”的效用小于“把这次抽奖的期望货币价值当作确定性的东西给你”的效用，即

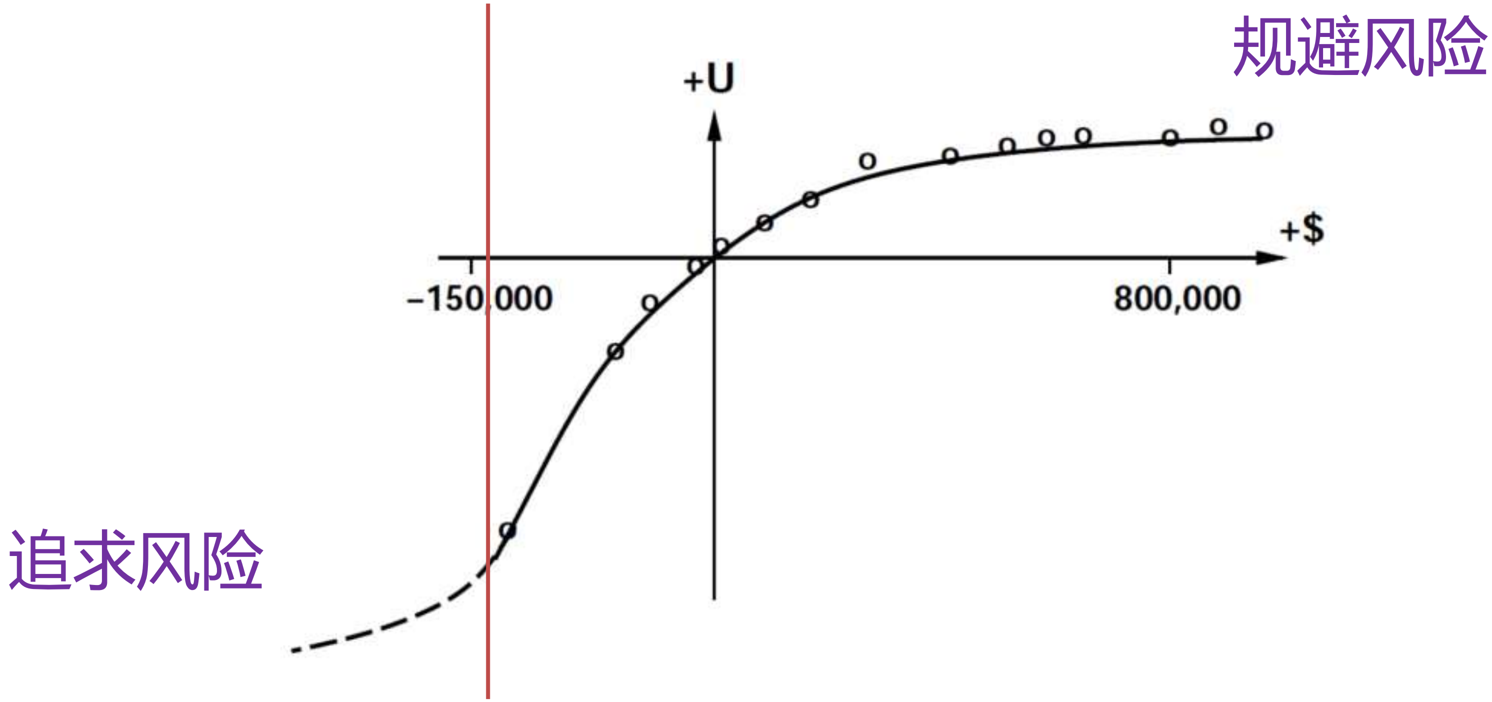
$$U(L) < U(S_{EMV(L)})$$



金钱的效用

- **风险中立**：对于任意抽奖L，“面对这次抽奖”的效用等于“把这次抽奖的期望货币价值当作确定性的东西给你”的效用，即

$$U(L) = U(S_{EMV(L)})$$

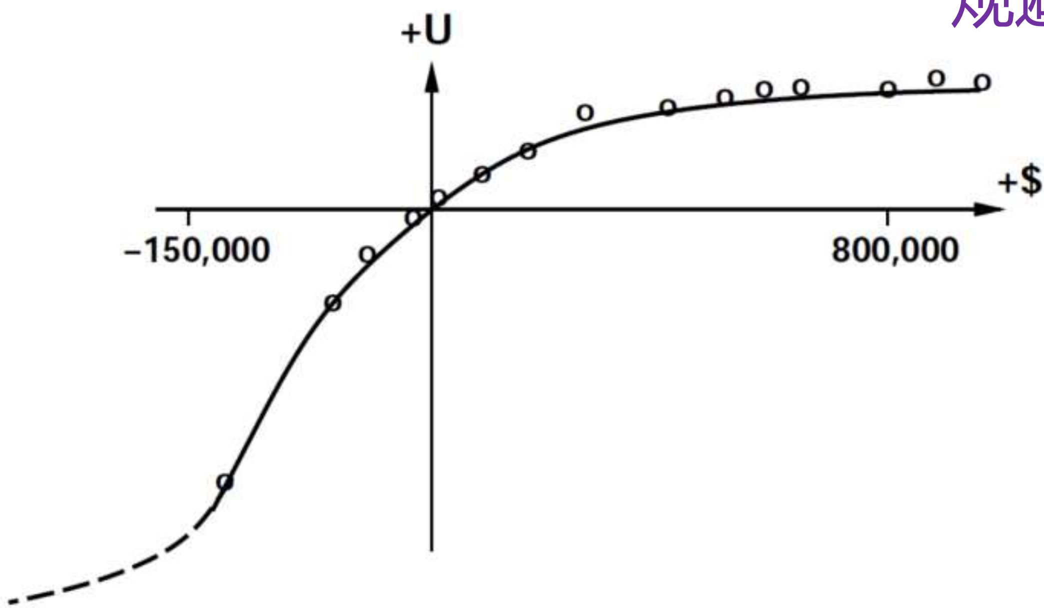


金钱的效用

- **保险费**：对于任意抽奖L，“面对这次抽奖”的效用减去“把这次抽奖的期望货币价值当作确定性的东西给你”的效用，即

$$U(L) - U(S_{EMV(L)})$$

规避风险



追求风险

主要内容

1. 理性偏好
2. 效用函数
3. 金钱的效用
4. 多属性效用函数

多属性效用函数

- 多属性效用理论：一个决策结果由两个或多个属性刻画
  - 发电厂有害物排放量：预防疾病，电力带来的好处等
  - 新机场的选址：环境污染，土地价格，飞机噪音，天气条件等

- 状态的效用通常依赖于一组属性$\mathbf{X} = X_1, \cdots, X_n$。
- 一个完整的赋值属性向量将是$\mathbf{x} = \langle x_1, \cdots, x_n \rangle$，其中$x_i$是一个数值或者是一个具有假设顺序的离散值。

- 例子：当建设新机场时，Deaths, Noise, Cost的效用是多少？

$$U(x_1, \cdots, x_n) = f[f_1(x_1), \cdots, f_n(x_n)]$$

多属性效用函数

- 理性Agent应该选择能够最大化期望效用的行为A：

$$Action = \underset{A}{\operatorname{argmax}} EU(A|E)$$

- 行为A的期望效用(Expected utility)定义为：

$$EU(A|E) = \sum_i P(Result_i(A)|Do(A), E)\, U(Result_i(A))$$

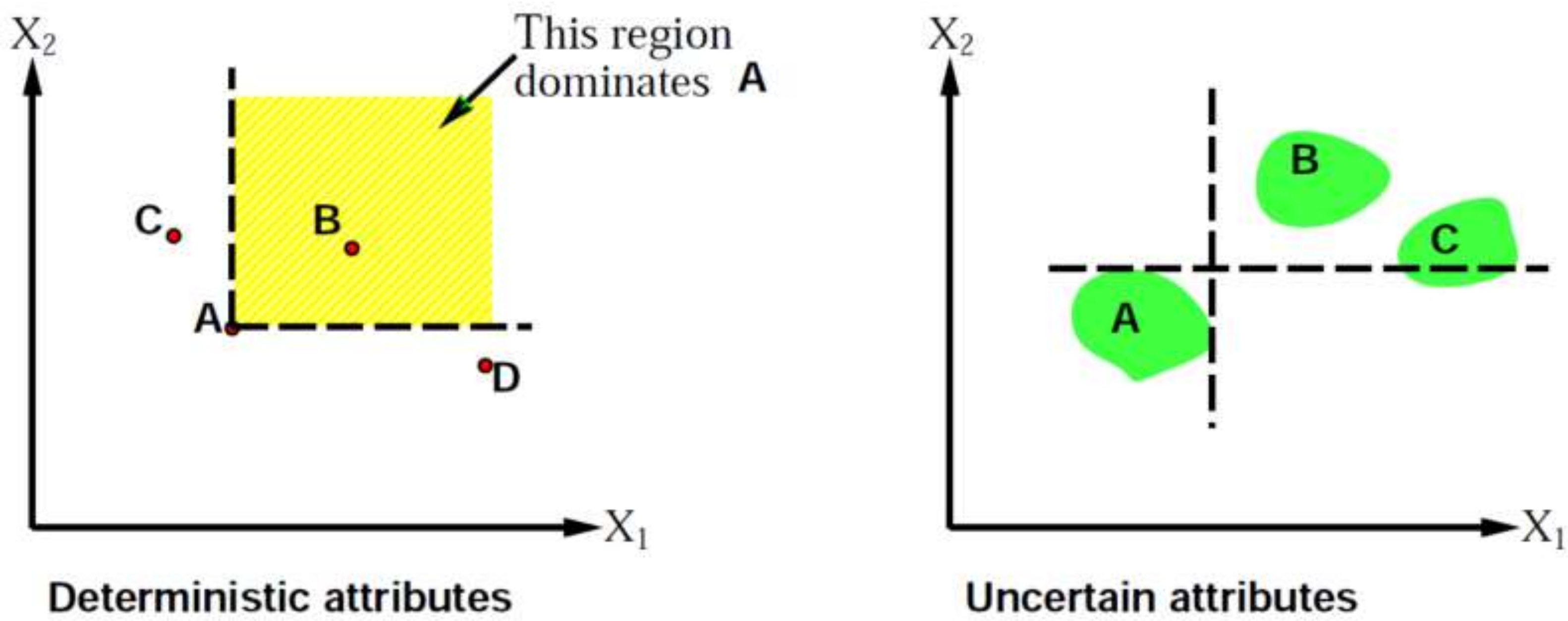
- 在确定性情况下，由行为A导致的状态是已知的：

$$EU(A) = U(State(x_1, \cdots, x_n))$$

严格优势

- 假设：更高的属性值对应更高的效用。
- 严格优势：行为B比行为A有严格优势，当且仅当

$$\forall i \quad X_i(B) \geq X_i(A) \qquad (\text{因此} EU(B) \geq EU(A))$$

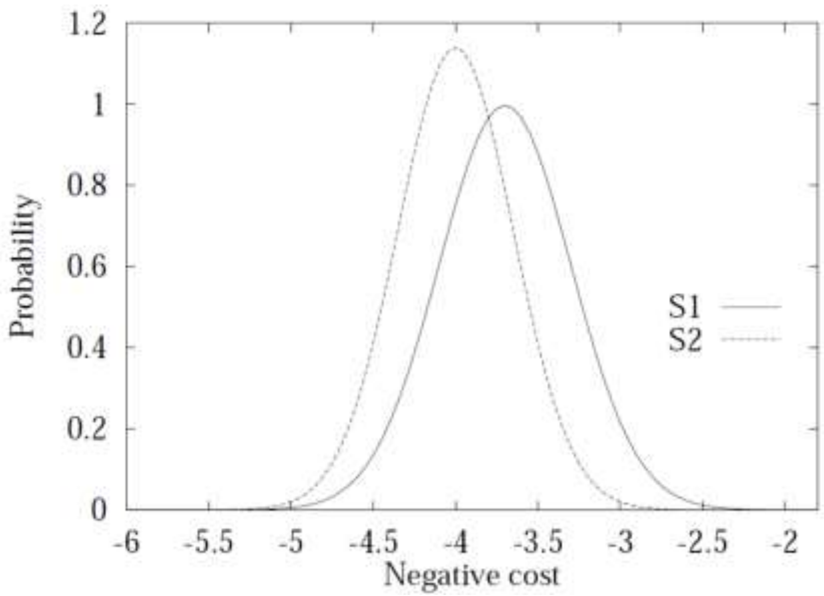


- 在现实中，很少存在严格优势。

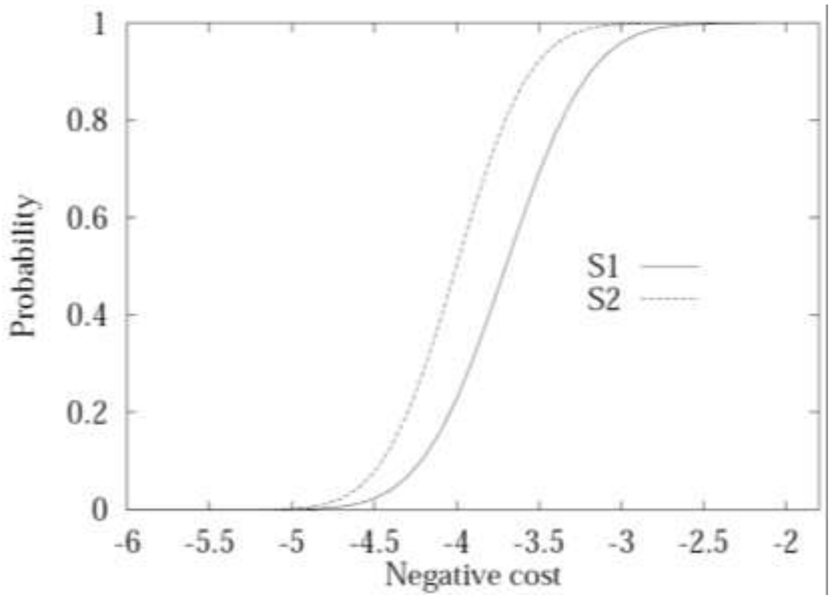
随机优势

- 如果两个行为A_1和A_2在属性X上导致概率分布$p_1(x)$和$p_2(x)$，那么当下式成立时，在X上A_1比A_2有随机优势：

$$\forall x \quad \int_{-\infty}^{x} p_1(x')\,dx' \le \int_{-\infty}^{x} p_2(x')\,dx'$$



机场建设费的负值



机场建设费的负值

随机优势

- 如果A_1比A_2有随机优势，那么对于任意单调非递减效用函数$U(x)$，A_1的期望效用大于等于A_2的期望效用，即

$$EU(A_1) \geq EU(A_2)$$

$$\int_{-\infty}^{+\infty} p_1(x)U(x)\,dx \geq \int_{-\infty}^{+\infty} p_2(x)U(x)\,dx$$

- 多属性案例：在所有属性上有随机优势 $\Rightarrow$ 最优的

随机优势

- 通常可以在没有精确分布的情况下，使用定性推理确定随机优势。
- 例如：
  - 交通费用随着货仓与城市间的距离增加而增大

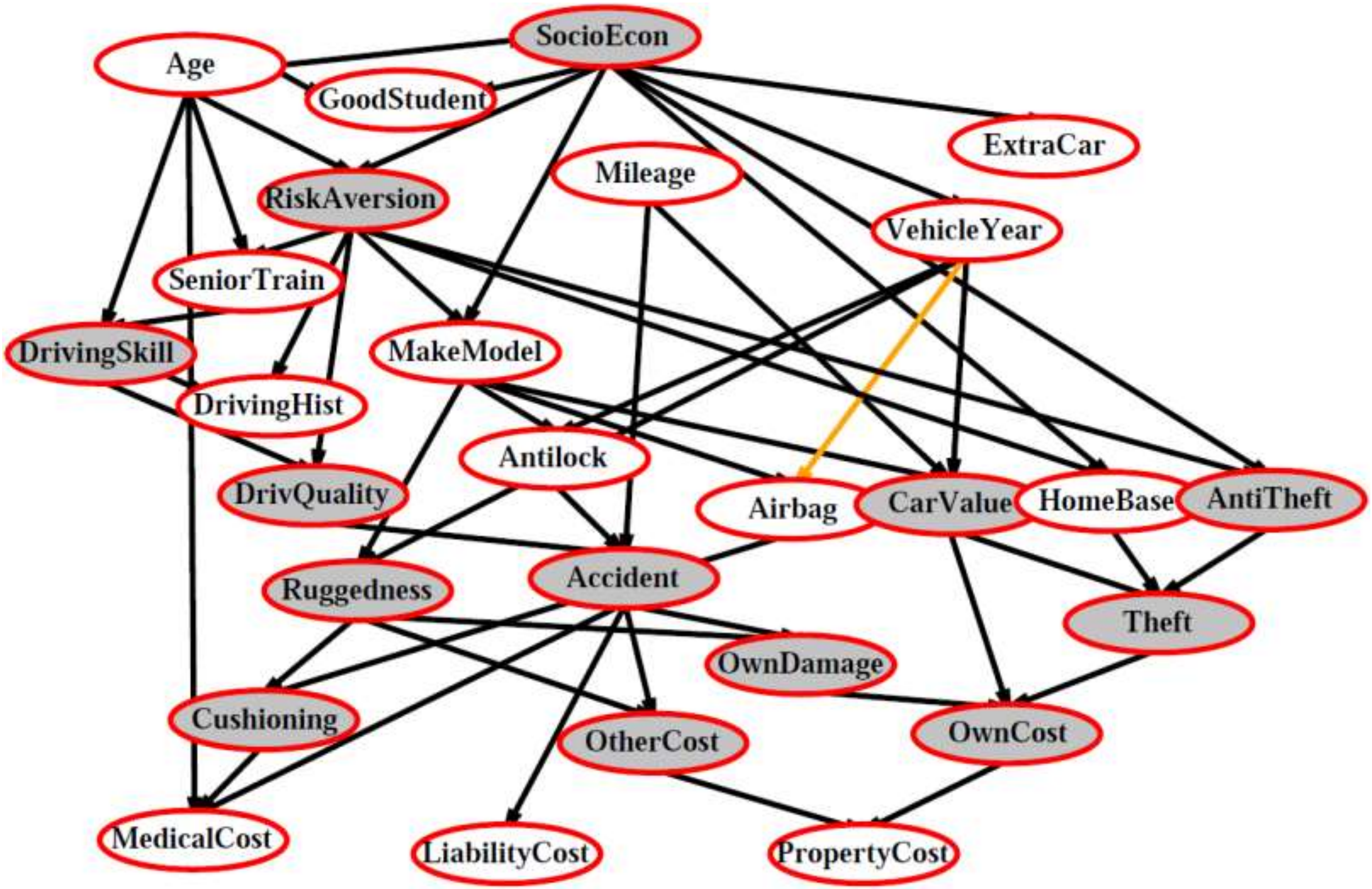
    S_1比S_2离城市近

    $\Rightarrow S_1$比S_2在交通费用上具有随机优势
  - 交通事故中的伤害随着碰撞速度增加而增大

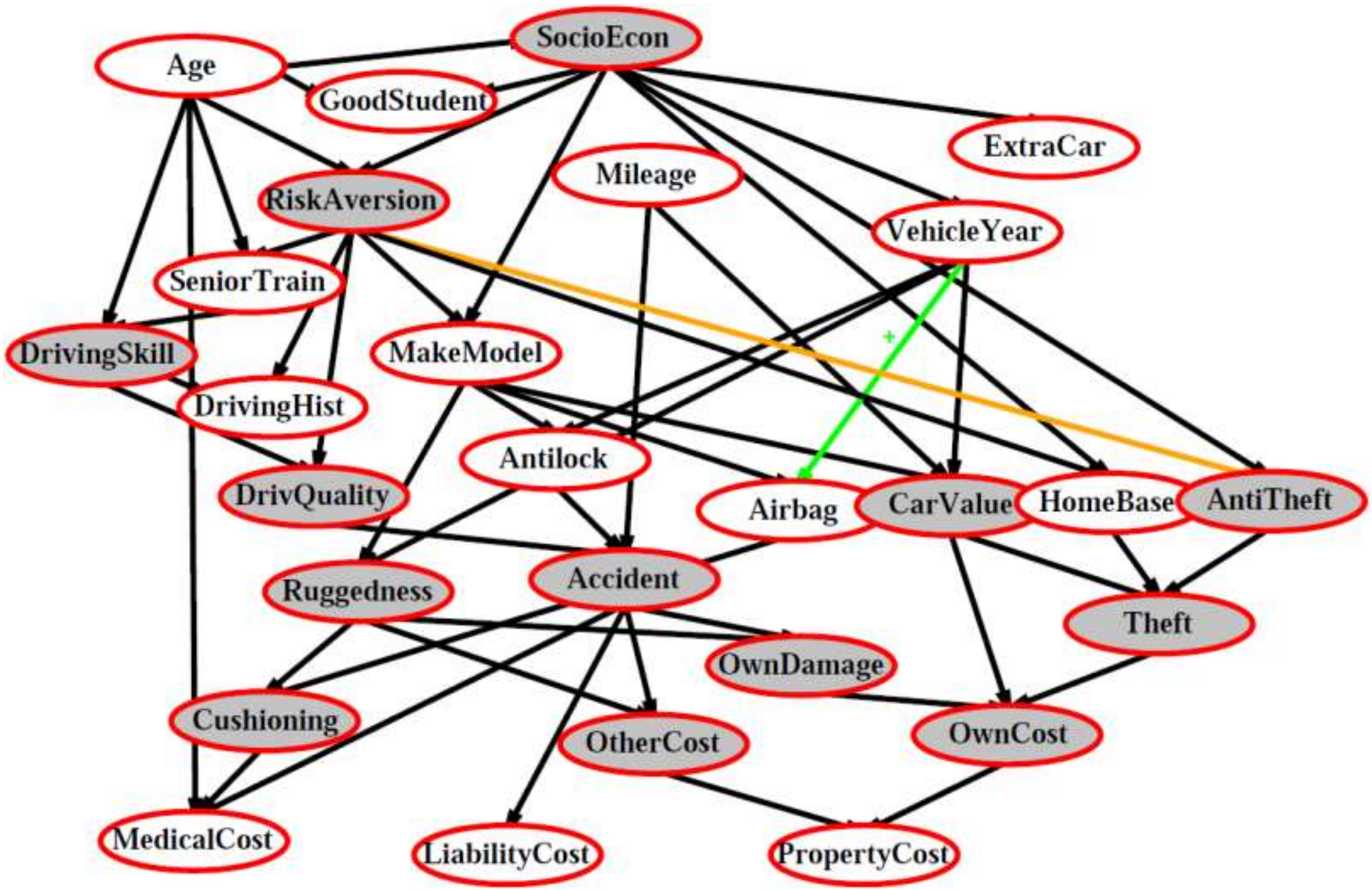
- 可以使用随机优势信息来标注信念网络：
  - $X \xrightarrow{+} Y$（X对Y有积极影响）
  - 对于Y的其他父母$\mathbf{Z}$的每个值$\mathbf{z}$

    $\forall x_1, x_2 \quad x_1 \geq x_2 \Rightarrow \mathrm{P}(Y|x_1, \mathbf{z})$比$\mathrm{P}(Y|x_2, \mathbf{z})$具有随机优势。

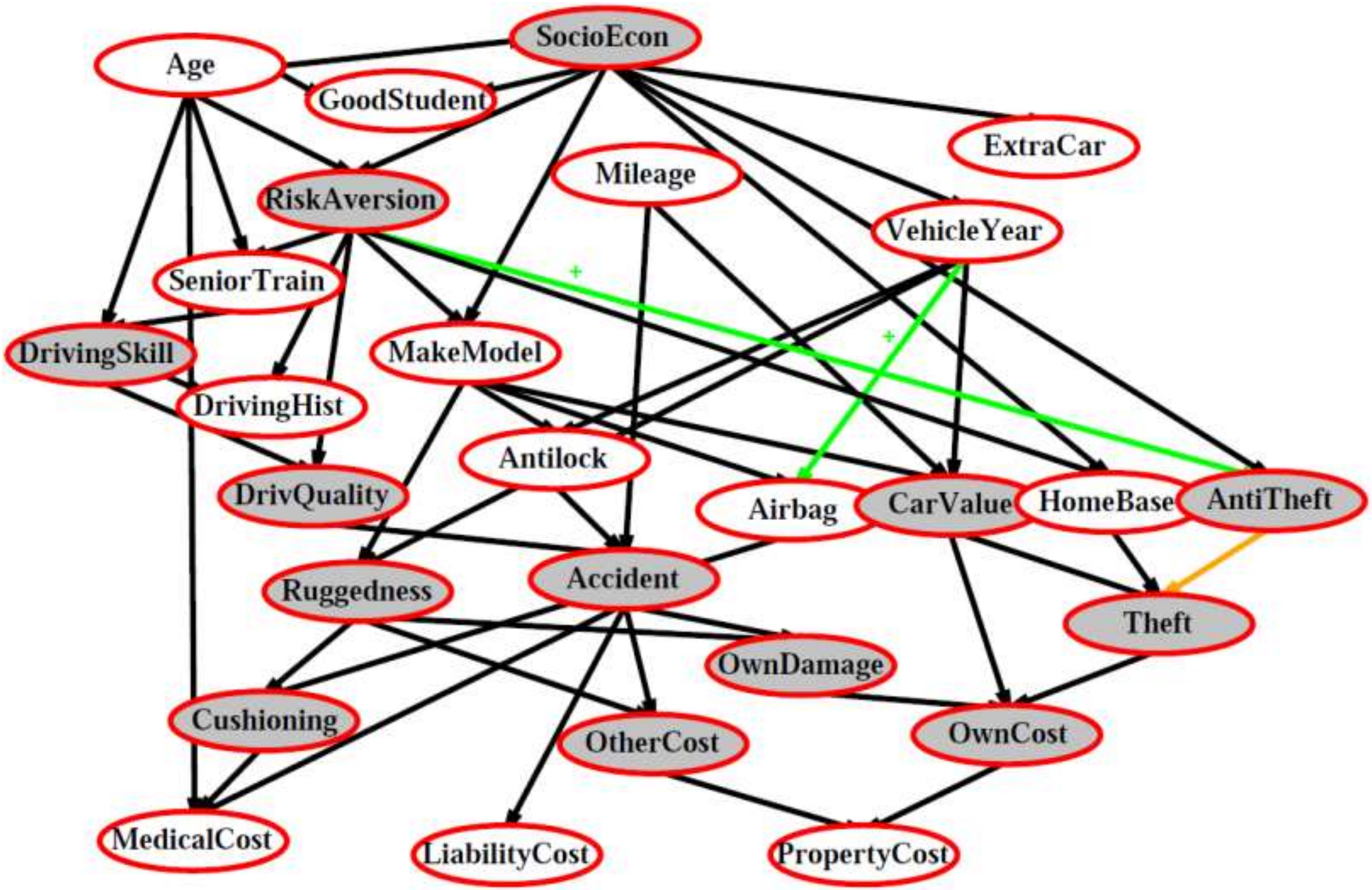
Label the arcs + or -



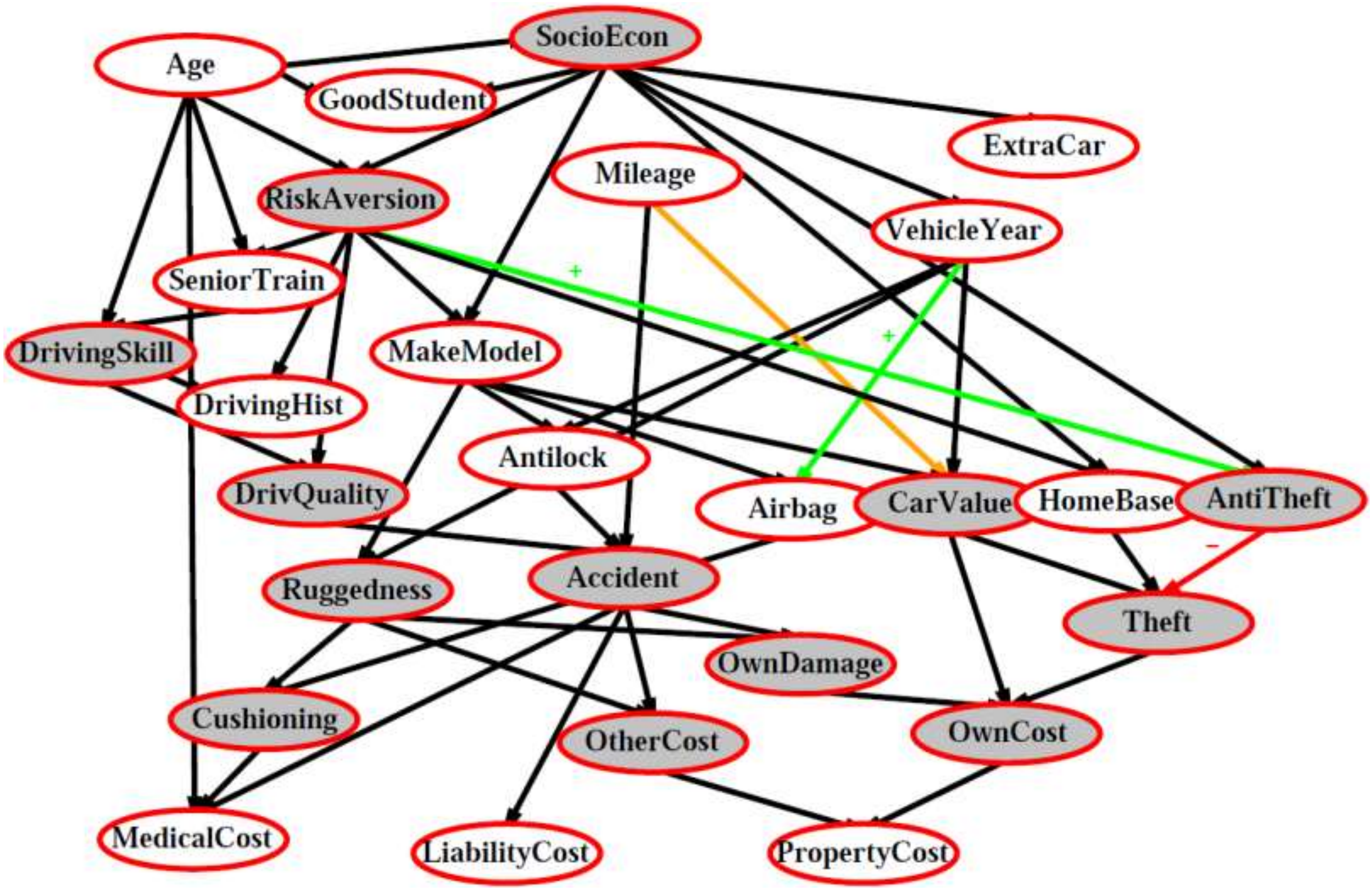
Label the arcs + or -



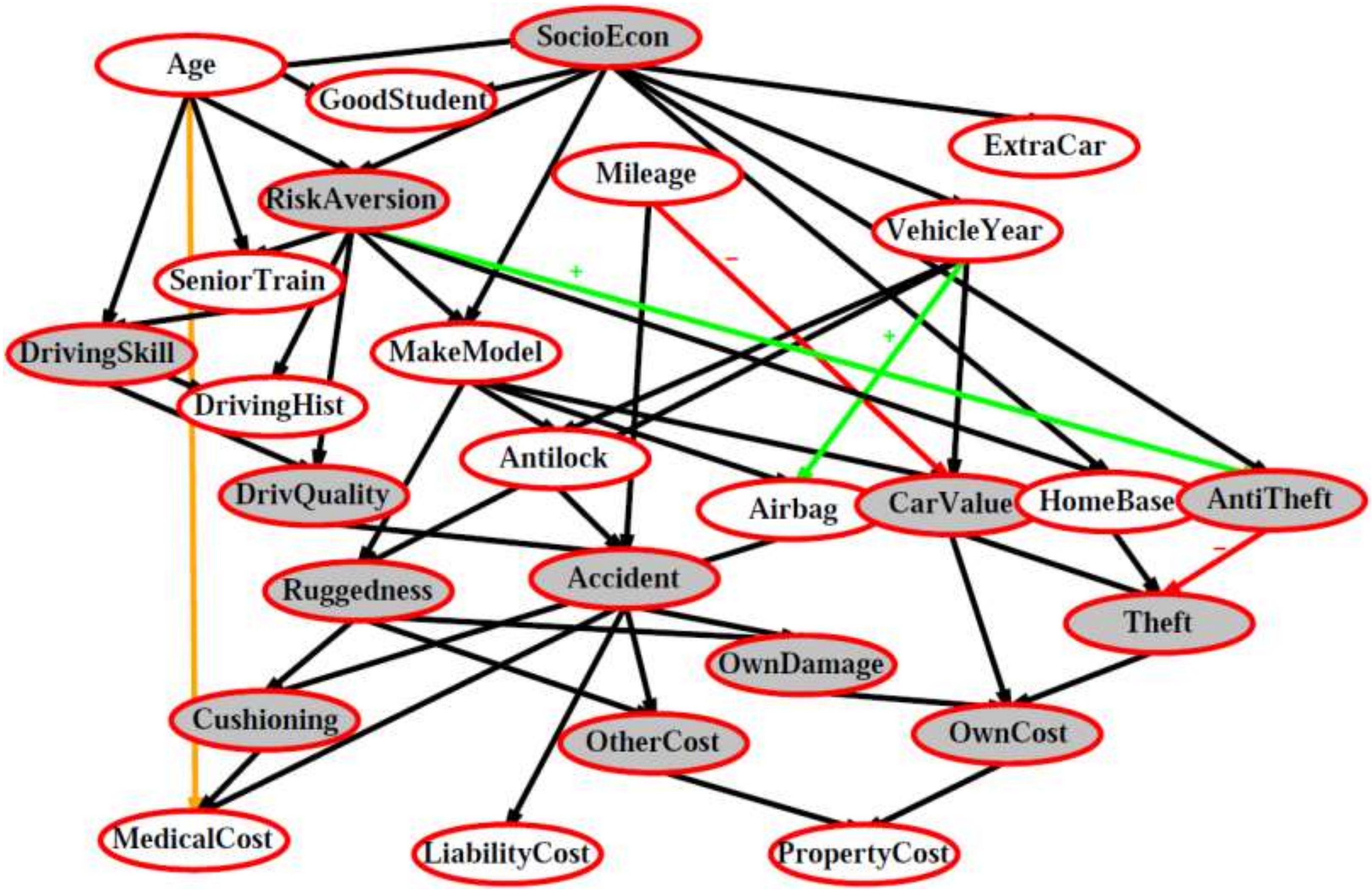
Label the arcs + or -



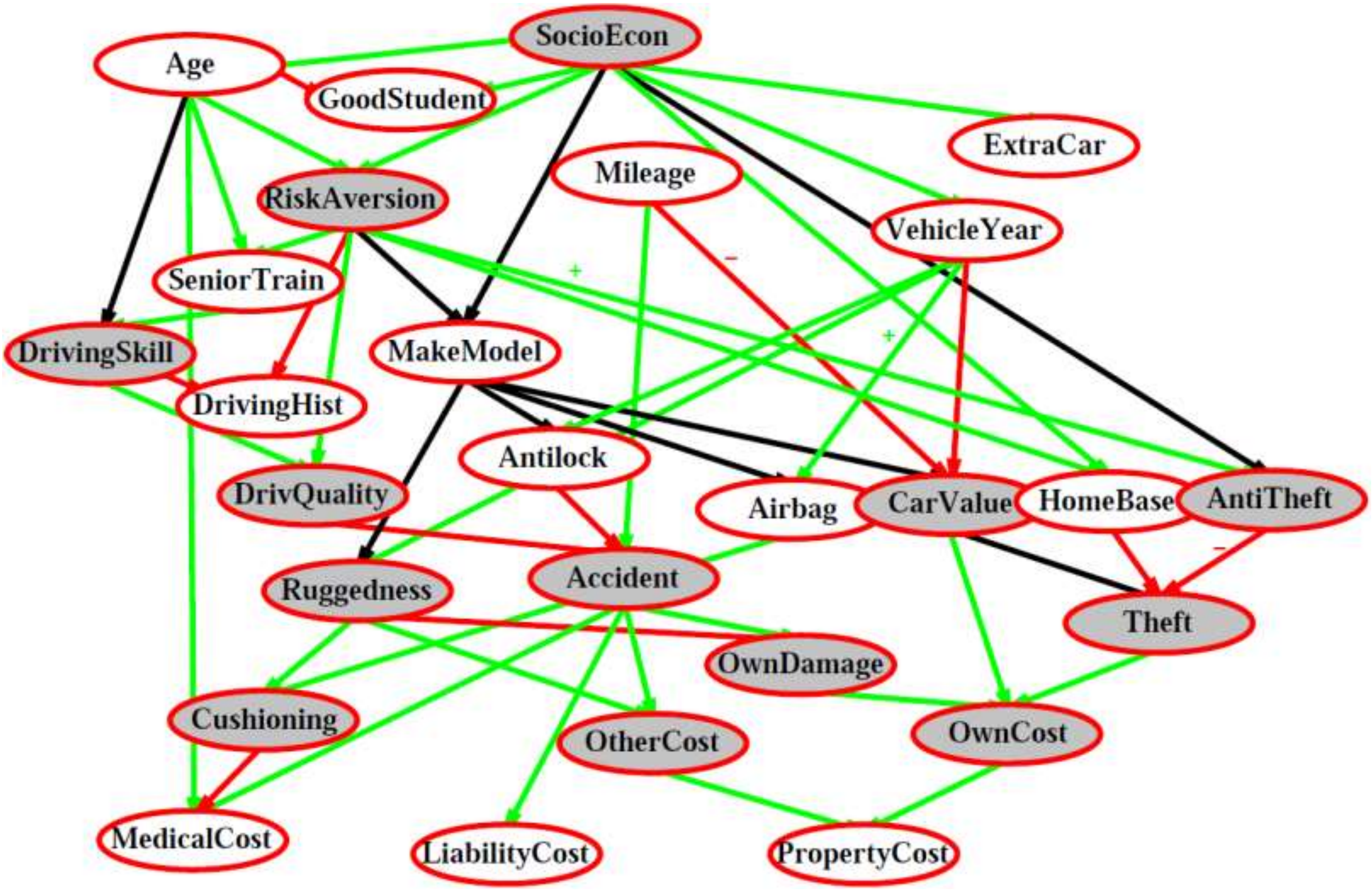
Label the arcs + or -



Label the arcs + or -



Label the arcs + or -



偏好结构：确定性

- 假设有n个属性$X_1, \cdots, X_n$，每个属性有m个可能值，那么共有m^n个状态。
- X_1和X_2偏好独立于X_3，当且仅当结果$\langle x_1, x_2, x_3 \rangle$和$\langle x_1', x_2', x_3 \rangle$之间的偏好不依赖于$x_3$。
- 例如：机场选址，$\langle Noise, Cost, Deaths \rangle$
  ① $U(\text{20,000 suffer}, \$4.6 \text{ billion}, 0.06 \text{ deaths/mpm}) \succ \langle \text{70,000 suffer}, \$4.2 \text{ billion}, 0.06 \text{ deaths/mpm} \rangle$
  ② $U(\text{20,000 suffer}, \$4.6 \text{ billion}, 0.05 \text{ deaths/mpm}) \succ \langle \text{70,000 suffer}, \$4.2 \text{ billion}, 0.05 \text{ deaths/mpm} \rangle$
  - 在两个城市的排名中，基于$Noise$和$Cost$不受$Deaths$的影响（两个城市的$Deaths$值相同）。
- 如果①和②所有取值都成立（其中$Deaths$值相同），那么$Noise$和$Cost$偏好独立于$Deaths$

偏好结构：确定性

- 相互偏好独立性：如果每个属性X_i都不会影响其它属性之间的权衡方式，那么所有属性显示出相互偏好独立性。
- 定理（Debreu，1960）：如果属性$X_1, \cdots, X_n$是相互偏好独立的，那么该Agent的偏好行为可被描述为最大化函数：

$$V(x_1, \cdots, x_n) = \sum_i V_i(x_i)$$

  其中，每个V_i是只涉及到属性X_i的一个值函数。
- 通常情况下采用线性（加法）值函数：

$$V(x_1, \cdots, x_n) = a_1 x_1 + \cdots + a_n x_n$$

  - 机场：$V(noise, cost, deaths) = -noise \times 10^4 - cost - deaths \times 10^{12}$

- 相互偏好独立性不成立的情形：狗、鸡、笼子

本章小结

1. 理性偏好
2. 效用函数
3. 金钱的效用
4. 多属性效用函数