Advanced Microeconomics I: Lecture 4

1 Non-expected utility framework

1 Subjective probability theory.

- Savage
- Anscombe and Aumann

2 Allais paradox

2 From Kahneman and Tversky (1979).

3 There are three possible monetary prizes: 2500, 2400, and 0.

4 Test 1: consider lotteries L_1 and L_1':

$$L_1 = (0.33, 0.66, 0.01) \qquad L_1' = (0, 1, 0).$$

82% choose L_1'.

5 Test 2: consider lotteries L_2 and L_2':

$$L_2 = (0.33, 0, 0.67) \qquad L_2' = (0, 0.34, 0.66).$$

83% choose L_2.

6 Paradox: Assume that there is a vNM expected utility function u.

- 82% choose L_1': $L_1' \succ L_1$, i.e.,

  $$0.33u(2500) + 0.66u(2400) + 0.01u(0) < u(2400).$$

  Thus,

  $$0.33u(2500) + 0.01u(0) < 0.34u(2400).$$

- 83% choose L_2: $L_2 \succ L_2'$, i.e.,

  $$0.33u(2500) + 0.67u(0) > 0.34u(2400) + 0.66u(0).$$

  Thus,

  $$0.33u(2500) + 0.01u(0) > 0.34u(2400).$$

7 Reaction 1 (by Marshack and Savage): choosing under uncertainty is a reflective activity in which one should be ready to correct mistakes if they are proven inconsistent with the basic principles of choice embodied in the independence axiom.

8 Reaction 2: Allais paradox is of limited significance for economics as a whole because it involves payoffs that are out of the ordinary and probabilities close to 0 and 1.

9 Reaction 3 (Regret theory): we could have $L_1 \succ L_1'$ because the expected regret caused by the possibility of getting zero in L_1' is too great.

10 Reaction 4: Give up the independence axiom.

3 Prospect theory

11 Prospect theory is a theory in cognitive psychology that describes the way people choose between probabilistic alternatives that involve risk, where the probabilities of outcomes are known.

The theory states that people make decisions based on the potential value of losses and gains rather than the final outcome, and that people evaluate these losses and gains using some heuristics.

The model is descriptive: it tries to model real-life choices, rather than optimal decisions, as normative models do.

The theory was created in 1979 and developed in 1992 by Daniel Kahneman and Amos Tversky as a psychologically more accurate description of decision making, compared to the expected utility theory.

12 Reference dependence 参照依赖.

People derive utility from gains and losses, measured relative to some reference point, rather than from absolute levels of wealth: the argument of $v(\cdot)$ is x_i, not $W + x_i$.

13 Loss aversion 损失规避.

People are much more sensitive to losses—even small losses—than to gains of the same magnitude.

14 Diminishing sensitivity 敏感度降低.

People tend to be risk-averse with respect to gains and risk-acceptant with respect to losses.

15 Probability weighting.

People do not weight outcomes by their objective probabilities p_i but rather by transformed probabilities or decision weights π_i.

The decision weights are computed with the help of a weighting function w whose argument is an objective probability.

Individuals overweight outcomes which are certain relative to outcomes which are merely probable.

They also overweight small probabilities and underweight moderate and high probabilities, and the latter effect is more pronounced than the former.

16 Model: Consider a gamble/lottery

$$(x_{-m}, p_{-m}; x_{-m+1}, p_{-m+1}; \ldots; x_0, p_0; \ldots; x_{n-1}, p_{n-1}; x_n, p_n),$$

where the notation should be read as "gain x_{-m} with probability p_{-m}, x_{-m+1} with probability p_{-m+1}, and so on," where the outcomes are arranged in increasing order, so that $x_i < x_j$ for $i < j$, and where $x_0 = 0$.

Under the expected utility theory, an individual evaluates the above gamble as

$$\sum_{i=-m}^{m} p_i u(W + x_i),$$

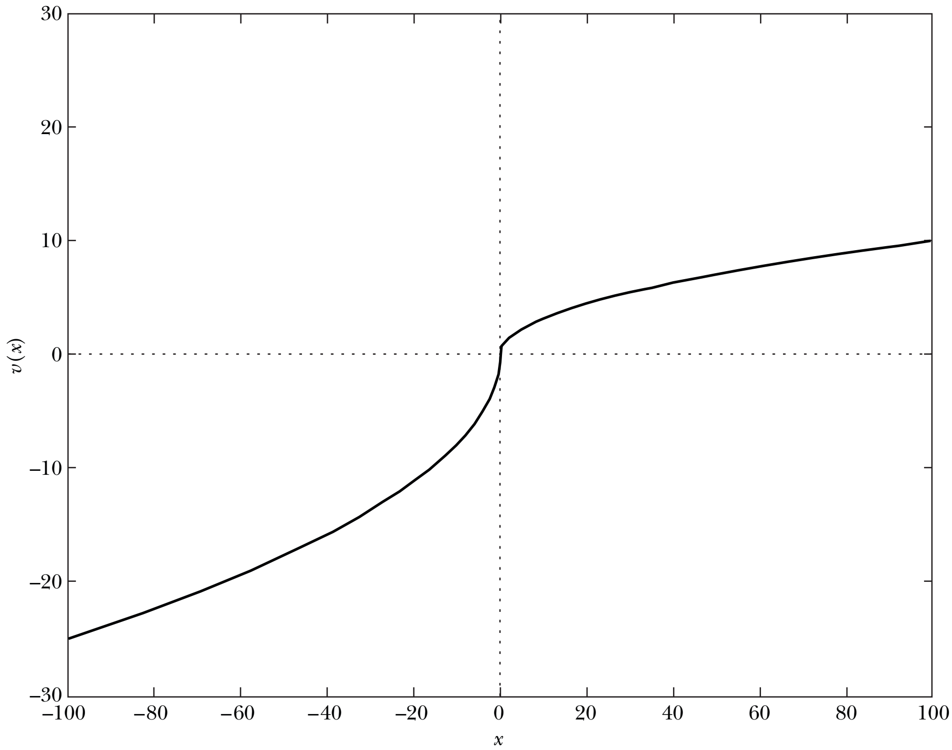
where W is current wealth and u is an increasing and concave vNM utility function.

Under the prospect theory, the gamble is evaluated as

$$\sum_{i=-m}^{m} w(p_i)v(x_i),$$

where v, the "value function," is an increasing function with $v(0) = 0$, and where $w(p_i)$ are "decision weights."

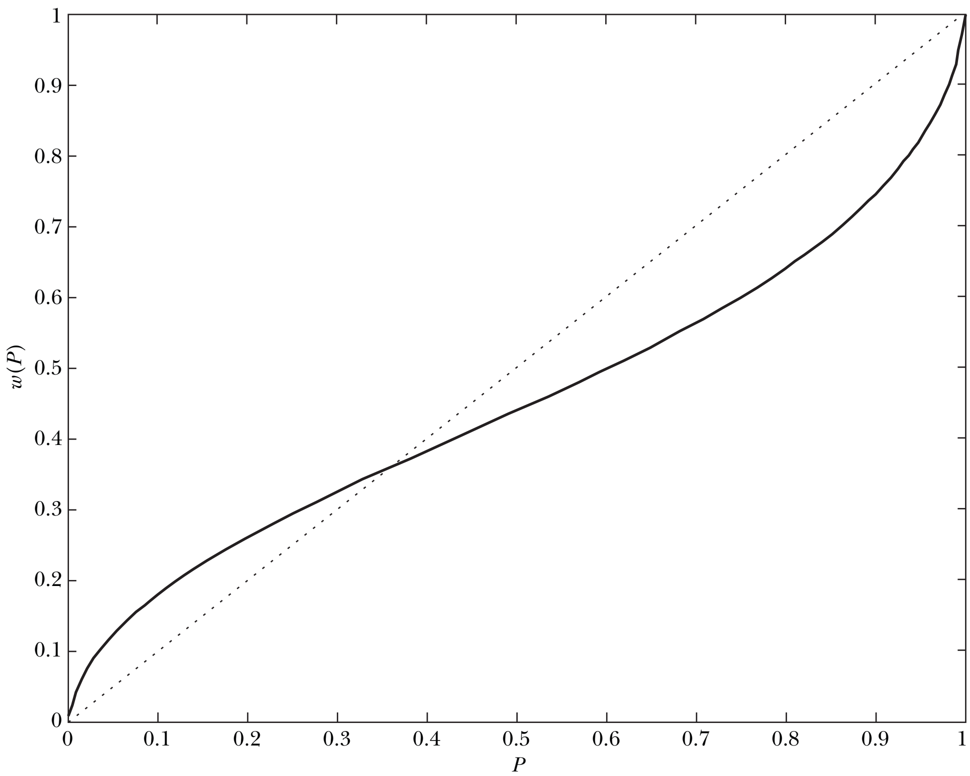
17 Graphs:



The graph plots the value function proposed by Tversky and Kahneman (1992) as part of prospect theory, namely $v(x) = x^\alpha$ for $x \geq 0$ and $v(x) = -\lambda(-x)^\alpha$ for $x < 0$, where x is a dollar gain or loss. The authors estimate $\alpha = 0.88$ and $\lambda = 2.25$ from experimental data. The plot uses $\alpha = 0.5$ and $\lambda = 2.5$ so as to make loss aversion and diminishing sensitivity easier to see.

- Loss aversion is generated by making the value function steeper in the region of losses than in the region of gains. Notice that the value placed on a \$100 gain, $v(100)$, is smaller in absolute magnitude than $v(-100)$, the value placed on a \$100 loss.

- The value function is concave in the region of gains but convex in the region of losses. while replacing a \$100 gain (or loss) with a \$200 gain (or loss) has a significant utility impact, replacing a \$1000 gain (or loss) with a \$1100 gain (or loss) has a smaller impact.



The graph plots the probability weighting function proposed by Tversky and Kahneman (1992) as part of prospect theory, namely $w(P) = \frac{P^\delta}{(P^\delta+(1-P)^\delta)^{1/\delta}}$, where P is an objective probability, for two values of δ. The solid line corresponds to $\delta = 0.65$, the value estimated by the authors from experimental data. The dotted line corresponds to $\delta = 1$, in other words, to linear probability weighting.

The weighting function overweights low probabilities and underweights high probabilities. People like both lotteries and insurance: they prefer a 0.001 chance of \$5000 to a certain gain of \$5, but also prefer a certain loss of \$5 to a 0.001 chance of losing \$5000.

18 The theory describes the decision processes in two stages:

- During an initial phase termed editing, outcomes of a decision are ordered according to a certain heuristic. In particular, people decide which outcomes they consider equivalent, set a reference point and then consider lesser outcomes as losses and greater ones as gains. The editing process can be viewed as composed of coding, combination, segregation, cancellation, simplification and detection of dominance.
- In the subsequent evaluation phase, people behave as if they would compute a value (utility), based on the potential outcomes and their respective probabilities, and then choose the alternative having a higher utility.

19 Revisit Allais paradox: Let $v(x) = x^{0.88}$ and $w(P) = \frac{P^\delta}{(P^\delta+(1-P)^\delta)^{1/\delta}}$ with $\delta = 0.65$.

4 Blackwell's theorem

20 个体在做决策时，虽然其效用（payoff）与现实世界的真实状态（state）有关，但往往无法观察到真实的状态。为了估计真实的状态，个体会考虑进行试验（experiment）以获取一些能够反应真实状态的信号（signal）。试验的好坏可以用其提供的信息量（或者更高的期望效用）来衡量。Blackwell 定理为试验之间的比较提供了建议一个简单的刻画。

21 Blackwell 定理由 David Blackwell 在 1951 年建立。值得一提的是，David Blackwell 是 UC Berkeley 第一个终身轨的黑人教授。

22 假设现实世界的真实状态有 n 种可能 $\{\omega_1, \omega_2, \ldots, \omega_n\}$，同时假设每个 ω_i 出现的概率是 p_i。记 $p = (p_1, p_2, \ldots, p_n)$。

23 假设试验 P 是一个 $n \times m$ 维的行随机矩阵（即每行的行和等于 1），其中 P_{ij} 表示当真实状态是 ω_i 时观察的信号是 s_j 的概率。不同的试验对应的可能的信号集合会不一样，因此维数 m 也可能不一样。

24 在观察到信号 s_j 之后，个体将在有限个选择 $A = \{a_1, a_2, \ldots, a_\ell\}$ 中进行选择。个体的效用函数依赖于自己的选择和真实的状态，可以用一个 $\ell \times n$ 维的矩阵 U 表示：U_{ki} 表示真实状态是 ω_i 时选择 a_k 的效用。

25 在决策问题中，个体的（混合）策略 D 是一个 $m \times \ell$ 维的行随机矩阵，其中 D_{jk} 表示个体观察到信号 s_j 时选择 a_k 的概率。

26 当真实状态是 ω_i 时，采用试验 P 和策略 D 得到的效用是

$$\sum_{j=1}^{m} P_{ij} \sum_{k=1}^{\ell} D_{jk} \cdot U_{ki} = (PDU)_{ii}$$

27 于是，$\mathrm{diag}(PDU) = \big((PDU)_{11}, (PDU)_{22}, \ldots, (PDU)_{nn}\big)$ 表示的是效用向量（payoff vector）。随着 D 变化，该效用向量也会改变，记所有可能的效用向量为 $B(P,U) = \{\mathrm{diag}(PDU) \mid D \text{ 是一个行随机矩阵}\}$。

28 定义：如果对于每个 U，$B(Q,U) \subseteq B(P,U)$，那么称试验 P 比试验 Q 拥有更多的信息量（more informative）。

29 因为真实状态的先验概率（prior）是 p，所以采用试验 P 和策略 D 得到的期望效用是

$$\sum_{i=1}^{n} p_i \sum_{j=1}^{m} P_{ij} \sum_{k=1}^{\ell} D_{jk} \cdot U_{ki} = \mathrm{trace}(PDU\hat{p})$$

这里的 $\hat{p}$ 是一个 $n \times n$ 的矩阵，(i,i) 位置元素是 p_i，非对角元素都是零。因此最大的期望效用为 $F(P,U,p) = \max_D \mathrm{trace}(PDU\hat{p})$。

30 Blackwell 定理：考虑两个试验 P（$n \times m$ 维）和 Q（$n \times m'$ 维），以下三条等价：

- 试验 P 比试验 Q 拥有更多的信息量，即对于每个 U，$B(Q,U) \subseteq B(P,U)$。
- 对于每个 U 和 p，$F(P,U,p) \geq F(Q,U,p)$。
- 存在一个行随机矩阵 M（合适的维数）使得 $Q = PM$。

31 "3 → 1" 和 "1 → 2" 是显然的。以下证明 "2 → 3"：

Proof. (1) 假定对于每个 $m \times m'$ 维的行随机矩阵 M，$Q \neq PM$。于是 $Q \notin E$，其中 $E = \{PM \mid M$ 是一个行随机矩阵$\}$。由于 E 是 $\mathbb{R}^{n \times m'}$ 中的凸闭子集。基于超平面分离定理，存在一个 $n \times m'$ 维的矩阵 G，使得对于每个 $m \times m'$ 维行随机矩阵 M，我们有

$$\text{trace}(G^t Q) > \text{trace}(G^t PM)$$

（这里需要指出的是 $n \times m'$ 维矩阵空间上的线性泛函都是 $\text{trace}(G^t \cdot)$ 的形式；可以参考stackexchange）

(2) 令 $U^t = \hat{p}^{-1} G$。所以，

$$\text{trace}(PDU\hat{p}) = \text{trace}(PDG^t) = \text{trace}(G^t PD) < \text{trace}(G^t Q) = \text{trace}(QU\hat{p})$$

(3) 因此我们有

$$\max_D \text{trace}(PDU\hat{p}) < \text{trace}(QU\hat{p}) \leq \max_D \text{trace}(QDU\hat{p})$$

矛盾！

□

32 简单来说，Blackwell 定理说明了，如果试验 P 比试验 Q 拥有的信息量更丰富，那么 $Q = PM$。这个矩阵 M 描述的是通过“篡改”试验 P 的结果来得到试验 Q 结果的过程，并且这一篡改过程与真实的状态毫无关系。由于矩阵 M 是一个行随机矩阵，所以通过试验 P 得到的后验概率（posterior）是通过试验 Q 得到的后验概率（posterior）的保留均值的伸展（mean-preserving spread），这一意味着前者承受的期望风险更小。

5 Homework

- Recommendation reading: *Thinking, Fast and Slow* by Kahneman.