How Reasonable are the Axioms?

All the axioms that Savage and von Neumann-Morgenstern use seem so reasonable.

- Savage views his axioms as characterizing rationality

Is that reasonable?

They certainly don't always characterize how people act . . .

Allais Paradox

Recall the *Allais paradox*:

The set of prizes is $X = \{\$0, \$1,000,000, \$5,000,000\}$.

- Which probability do you prefer:
  $p_1 = (0.00, 1.00, 0.00)$ or $p_2 = (0.01, 0.89, 0.10)$?

Allais Paradox

The set of prizes is $X = \{\$0, \$1,000,000, \$5,000,000\}$.

- Which probability do you prefer:
  $p_1 = (0.00, 1.00, 0.00)$ or $p_2 = (0.01, 0.89, 0.10)$?
- Which probability do you prefer:
  $p_3 = (0.90, 0.00, 0.10)$ or $p_4 = (0.89, 0.11, 0.00)$?

Many subjects report: $p_1 \succ p_2$ and $p_3 \succ p_4$

Inconsistent with Maximizing Expected Utility

Suppose (u_0, u_1, u_5) represents $\succ$.
Then $p_1 \succ p_2$ implies

$$\begin{aligned} u_1 &> .01u_0 + .89u_1 + .1u_5 \\ .11u_1 - .01u_0 &> .1u_5 \\ .11u_1 + .89u_0 &> .1u_5 + .9u_0. \end{aligned}$$

So $p_4 \succ p_3$.
Which axiom is violated?
Independence: $a \succ b$ iff $\alpha a + (1-\alpha)c \succ \alpha b + (1-\alpha)c$.

- homework – explain exactly how.

Ellsberg Paradox

We considered the *Ellsberg paradox* in the first class:

There is one urn with 90 balls: 30 of these balls are red (R) and the rest are either blue (B) or yellow (Y). Consider the following two choice situations:

I: a. Win $\$100$ if a ball drawn from the urn is R and nothing otherwise.
a'. Win $\$100$ if a ball drawn from the urn is B and nothing otherwise.

Ellsberg Paradox

There is one urn with with 90 balls: 30 of these balls are red (R) and the rest are either blue (B) or yellow (Y). Consider the following two choice situations:

I:

- a. Win $\$100$ if a ball drawn from the urn is R and nothing otherwise.
- a'. Win $\$100$ if a ball drawn from the urn is B and nothing otherwise.

II:

- b. Win $\$100$ if a ball drawn from the urn is R or Y and nothing otherwise.
- b'. Win $\$100$ if a ball drawn from the urn is B or Y and nothing otherwise.

Inconsistent with SEU

Suppose a decision maker's preferences are such that $a \succ a'$ and $b' \succ b$.
If there are subjective probabilities then the first choice implies that the probability of a red ball is greater than the probability of a blue ball and the second choice implies the reverse.

Which of Savage's axioms is violated?

- Independence: Remember that an act is a function from states to outcomes. Let $T \subseteq S$ be a subset of states. Then

$$f_T g \succeq f'_T g \text{ iff } f_T h \succeq f'_T h.$$

Homework: prove that the standard choices in the Ellsberg paradox violate this.

These examples suggest that maximizing expected utility is not obviously always the “right” thing to do.

- But if we don’t do that, what should we do?
  - We’ve already seen some alternatives (regret, maximin, . . .)
  - Let’s go back to one of them: maxmin expected utility

Maxmin Expected Utility Rule

Suppose that the decision maker's uncertainty can be represented by a set $\mathcal{P}$ of probabilities . Let

$$\underline{E}_{\mathcal{P}}(u_a) = \inf_{\Pr \in \mathcal{P}} \{E_{\Pr}(u_a) : \Pr \in \mathcal{P}\}$$

Recall the maximin expected utility rule: (covered earlier in the course):

- $a >^1_{\mathcal{P}} a'$ iff $\underline{E}_{\mathcal{P}}(u_a) > \underline{E}_{\mathcal{P}}(u_{a'})$

This is like maximin:

- Optimizing the worst-case expectation

This could explain the Ellsberg Paradox:

- Let $\mathcal{P} = \{(1/3, p_B, p_Y) : 0 \leq p_B \leq 2/3\}$

Gilboa and Schmeidler axiomatized the maxmin expected utility rule

- It does *not* satisfy independence
- Gilboa and Schmeidler replaced independence by a weaker axiom.

Why independence may not be so reasonable

- Suppose that there are four states, w, x, y, z.
- f is the act where you get 1 in state w, 0 in all other states
- f' is the act where you get 1 in state x, 0 in all other states
- g is the act where you get 1 in state y, 0 in all other states
- h is the act where you get 1 in state z, 0 in all other states
- Let $A = \{w, x\}$.

Why independence may not be so reasonable

- Suppose that there are four states, w, x, y, z.
- f is the act where you get 1 in state w, 0 in all other states
- f' is the act where you get 1 in state x, 0 in all other states
- g is the act where you get 1 in state y, 0 in all other states
- h is the act where you get 1 in state z, 0 in all other states
- Let $A = \{w, x\}$.

act	w	x	y	z
$f_A g$	1	0	1	0
$f'_A g$	0	1	1	0
$f_A h$	1	0	0	1
$f'_A h$	0	1	0	1

Why independence may not be so reasonable

- Suppose that there are four states, w, x, y, z.
- f is the act where you get 1 in state w, 0 in all other states
- f' is the act where you get 1 in state x, 0 in all other states
- g is the act where you get 1 in state y, 0 in all other states
- h is the act where you get 1 in state z, 0 in all other states
- Let $A = \{w, x\}$.

act	w	x	y	z
$f_A g$	1	0	1	0
$f'_A g$	0	1	1	0
$f_A h$	1	0	0	1
$f'_A h$	0	1	0	1

Now suppose that $\Pr(\{w, y\}) = \Pr(\{x, z\}) = 1/2$, but you don't know the probability of individual states. It seems reasonable that

- $f_A g \succ f'_A g$ (getting 1 with probability $1/2$ is better than getting 1 with some unknown probability)
- $f'_A h \succ f_A h$ (same logic).

Why independence may not be so reasonable

- Suppose that there are four states, w, x, y, z.
- f is the act where you get 1 in state w, 0 in all other states
- f' is the act where you get 1 in state x, 0 in all other states
- g is the act where you get 1 in state y, 0 in all other states
- h is the act where you get 1 in state z, 0 in all other states
- Let $A = \{w, x\}$.

act	w	x	y	z
$f_A g$	1	0	1	0
$f'_A g$	0	1	1	0
$f_A h$	1	0	0	1
$f'_A h$	0	1	0	1

Now suppose that $\Pr(\{w, y\}) = \Pr(\{x, z\}) = 1/2$, but you don't know the probability of individual states. It seems reasonable that

- $f_A g \succ f'_A g$ (getting 1 with probability $1/2$ is better than getting 1 with some unknown probability)
- $f'_A h \succ f_A h$ (same logic).

Independence fails!

Comonotonic independence

The weaker independent axiom used by Gilboa and Schmeidler is called *comonotonic independence*:

Comonotonic independence

The weaker independent axiom used by Gilboa and Schmeidler is called *comonotonic independence*:

Acts f and g are *comonotonic* if there do not exist states s and t such that

$$f(s) \succ f(t) \text{ and } g(t) \succ g(s)$$

- f and g are comonotonic if you can't be happier to be in state s than state t when doing f and be happier to be in state t than state s when doing g.
- If h is a constant act, then f and h are comonotonic for all acts f (since we never have $h(s) \succ h(t)$).

Comonotonic independence

The weaker independent axiom used by Gilboa and Schmeidler is called *comonotonic independence*:

Acts f and g are *comonotonic* if there do not exist states s and t such that

$$f(s) \succ f(t) \text{ and } g(t) \succ g(s)$$

- f and g are comonotonic if you can't be happier to be in state s than state t when doing f and be happier to be in state t than state s when doing g.
- If h is a constant act, then f and h are comonotonic for all acts f (since we never have $h(s) \succ h(t)$).
- [Comonotonic Independence:] If f and h and g and h are both comonotonic, then $f \succ g$ iff for all $\alpha \in (0, 1]$, $\alpha f + (1 - \alpha)h \succ \alpha g + (1 - \alpha)h$.

Gilboa and Schmeidler proved a representation theorem for their axiomatization:

- An agent's preference order obeys the Gilboa-Schmeidler axioms iff he has a utility function (unique up to affine transformations) and a set $\mathcal{P}$ of probability measures such that $a_1 \succ a_2$ iff $a_1 >^1_{\mathcal{P}} a_2$.
  - If the agent's preference order obeys the axioms iff he is acting like a maxmin expected utility maximizer
  - $\mathcal{P}$ is not unique, but its convex closure is.
  - The *convex closure* of $\mathcal{P}$ is the smallest closed convex set containing $\mathcal{P}$.
    - If $p, p' \in P$, then $\alpha p + (1-\alpha)p' \in \mathcal{P}$
    - $\mathcal{P}$ also contains its limit points
  - if $(u, \mathcal{P})$ and $(u, \mathcal{P}')$ both represent the axioms, then the convex closure of $\mathcal{P}$ = the convex closure of $\mathcal{P}'$

Gilboa and Schmeidler proved a representation theorem for their axiomatization:

- An agent's preference order obeys the Gilboa-Schmeidler axioms iff he has a utility function (unique up to affine transformations) and a set $\mathcal{P}$ of probability measures such that $a_1 \succ a_2$ iff $a_1 >^1_{\mathcal{P}} a_2$.
  - If the agent's preference order obeys the axioms iff he is acting like a maxmin expected utility maximizer
  - $\mathcal{P}$ is not unique, but its convex closure is.
  - The *convex closure* of $\mathcal{P}$ is the smallest closed convex set containing $\mathcal{P}$.
    - If $p, p' \in P$, then $\alpha p + (1-\alpha)p' \in \mathcal{P}$
    - $\mathcal{P}$ also contains its limit points
  - if $(u, \mathcal{P})$ and $(u, \mathcal{P}')$ both represent the axioms, then the convex closure of $\mathcal{P}$ $=$ the convex closure of $\mathcal{P}'$

While maxmin expected utility captures some aspects of human behavior, but certainly doesn't capture all its complexities.

Cumulative Prospect Theory

Cumulative prospect theory (CPT) is a model of decision making due to Tversky and Kahneman (1992)

- An improvement over *prospect theory*, their earlier approach
- CPT was a significant part of why Kahneman won the Nobel Prize (Tversky would have won it too had he not died)
- It is largely viewed as the best descriptive theory of decision making that we have
  - It's used by practitioners

Key insights of CPT:

- People care more about losses than gains
- People tend to overweight extreme unlikely events and underweight "average" events.

CPT: Technical details

In the utility functions considered in homework and in class up to now, we took the utility of (final) wealth

- CPT considers a reference point, and takes utility of wealth relative to that reference point
  - Above the reference point is a gain; below it is a loss.

- The DM is choosing among *prospects* (i.e., acts) mapping a state space S to a finite outcome space $X = \{x_{-m} \ldots, x_{-1}, x_0, \ldots, x_n\}$.
  - The x_i's are ordered by goodness: $x_j \succ x_i$ if $j > i$
  - x_0 is the reference point
  - $v(x_0) = 0$; $v(x_i) > 0$ if $i > 0$; $v(x_i) < 0$ if $i < 0$.
- For prospect f, A_i^f is the set of states s such that $f(s) = x_i$.
- $V(f) = \sum_{i=-m}^{n} \pi_i^f v(x_i)$.
- Like expected utility, but π_i^f is not the probability of getting outcome i with act f

- The DM is choosing among *prospects* (i.e., acts) mapping a state space S to a finite outcome space $X = \{x_{-m} \dots, x_{-1}, x_0, \dots, x_n\}$.
  - The x_i's are ordered by goodness: $x_j \succ x_i$ if $j > i$
  - x_0 is the reference point
  - $v(x_0) = 0$; $v(x_i) > 0$ if $i > 0$; $v(x_i) < 0$ if $i < 0$.
- For prospect f, A_i^f is the set of states s such that $f(s) = x_i$.
- $V(f) = \sum_{i=-m}^{n} \pi_i^f v(x_i)$.
- Like expected utility, but π_i^f is not the probability of getting outcome i with act f
- A *capacity* W is a generalization of a probability measure:
  - $W(\emptyset) = 0$; $W(S) = 1$; $W(A) \leq W(B)$ if $A \subseteq B$.
- $\pi_n^f = W(A_n^f)$; $\pi_{-m}^f = W(A_{-m}^f)$
- $\pi_i^f = W(A_i^f \cup \dots \cup A_n^f) - W(A_{i+1}^f \cup \dots \cup A_n^f)$ if $0 \leq i \leq n-1$
- $\pi_i^f = W(A_{-m}^f \cup \dots \cup A_i^f) - W(A_{-m}^f \cup \dots \cup A_{i-1}^f)$ if $-m < i \leq 0$
  - W lets us treat losses differently from gains
  - W lets us overweight extreme unlikely events (A_{-m}^f and A_n^f)
  - if W is a probability, then π_i^f is the probability of getting outcome x_i with f.

Framing Effects—Kahneman and Tversky

A disease is expected to kill 600 people. Two alternative programs have been proposed:

- Program A: 200 people will be saved
- Program B:
  - probability 1/3: 600 people will be saved
  - probability 2/3: no one will be saved

Which program would you favor?

Framing Effects—Kahneman and Tversky

A disease is expected to kill 600 people. Two alternative programs have been proposed:

- Program C: 400 people will die
- Program D:
  - probability 1/3: no one will die
  - probability 2/3: 600 will die

Which program would you favor?

Framing Effects—Kahneman and Tversky

Kahneman and Tversky found:

- 72% chose A over B.
- 22% chose C over D.

But if 200 people will be saved out of 600 is the same to the decision-maker as 400 people will die out of 600, and so on, then A and C are identical and so are B and D.

- losses (death) are weighed more heavily than gains (people staying alive)

Conjunction Fallacy or Failure of Extensionality

Tom is a rancher from Montana.
Which bet would you prefer?

- Win $\$10$ if Tom drives either a Ford or a Chevy, otherwise win nothing
- Win $\$10$ if Tom drives either a Chevy truck or Ford truck, otherwise win nothing

Kahneman and Tversky experiment:
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

85% of subjects chose the second option.

Kahneman and Tversky experiment:
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

Another systematic error: Ignoring priors

A cab was involved in a hit and run accident last night. Two cab companies, Green and Blue, operate in the city.
You know:

- A witness identified the cab as Blue.
- Witnesses are pretty reliable: Tests have shown that in similar cirumstances witnesses correctly identify each of the two cabs 80% of the time and misidentify them 20% of the time.
- 85% of the cabs in the city are Green the rest are Blue.

What is the probability that the cab involved in the accident was Blue?

The correct answer requires Bayes rule:

$$\begin{aligned} Pr(B|idB) &= \frac{Pr(idB|B)Pr(B)}{Pr(idB)} \\ &= \frac{(.8)(.15)}{(.8)(.15) + (.2)(.85)} \\ &= .41 \end{aligned}$$

Computational limitations

People use "fast and frugal" heuristics [Gigerenzer]

- simple decision rules for making decisions that often work surprisingly well
- Which has a larger population: Detroit or Milwaukee?
  - Europeans do better on this question than Americans!

Computational limitations

People use "fast and frugal" heuristics [Gigerenzer]

- simple decision rules for making decisions that often work surprisingly well
- Which has a larger population: Detroit or Milwaukee?
  - Europeans do better on this question than Americans!

Computational limitations

People use “fast and frugal” heuristics [Gigerenzer]

- simple decision rules for making decisions that often work surprisingly well
- Which has a larger population: Detroit or Milwaukee?
  - Europeans do better on this question than Americans!

Claim: People are rational, but the “mistakes” that we’re seeing are the outcome of computational limitations [Wilson; Halpern/Pass/Seeman]