More on learning / Rationality and choice

Minds & Machines
SymSys 1/Phil 99/Psych 35/Ling 35
11/1/16
Homework

Focus on Bayesian inference
Straightforward but not short
Get an early start!

Probability bonus section
Today, 3-4, 200-002
Continuing with learning …
Rationality & choice
What's the point of having a brain?

The fundamental problem of cognition: making good choices under conditions of uncertainty
The B-D-I model of action

belief → intention → desire → action → goal

Bratman, 1987
Some concepts of rationality

What does it mean to be ‘rational’?

1. Having beliefs justified by your evidence
2. Having coherent beliefs
3. Having the ‘right’ values (?)
4. Having coherent values
5. Making choices that could be justified by some coherent beliefs & values
Bayesian norms of rational belief

1. A rational agent’s beliefs must be **probabilistically coherent** – representable by a probability measure $P$.

2. A rational agent’s beliefs must **change in response to evidence** and must do so by **conditioning** $P$ on the evidence.
Norms of rational choice

The basic Bayesian model:

A rational agent’s choices must

**maximize expected utility**

i.e., emerge as the best options relative to a
the agent’s desires and (probabilistic) beliefs.
Rational choice theory
developed out of theory of gambling
- should I gamble at all?
- If I do, when should I stop and go home?
- what’s more profitable: roulette or blackjack?
- how much can I expect to make on this bet?
- Is the potential gain worth the risk of losing?

artificial?
- everything you do is a gamble!
Norms of rational choice

Suppose you prefer eating

squash to peas,  
peas to carrots,  
carrots to rice.

Can you also prefer eating rice to squash?

Transitivity of preference

If you prefer A to B and B to C,  
then you prefer A to C.
Norms of rational choice

Suppose you prefer squash to peas.
Can you also prefer peas to squash?

**Irreflexivity of preference**

If you prefer A to B, then you do not prefer B to A.
Norms of rational choice

Suppose you do not prefer squash to peas, and you also don’t prefer peas to squash. Does it follow that you are indifferent?

Completeness of preference

For any A and B, you either

1. prefer A to B;
2. prefer B to A; or
3. are indifferent between them.
Norms of rational choice

If you prefer \textbf{squash} to \textbf{peas}, can you prefer
– a 50/50 chance of squash vs. carrots
to
– a 50/50 chance of peas vs. carrots?

\textbf{Independence of irrelevant alternatives:}
If you prefer \textbf{A} to \textbf{B}, then you prefer
a n\% chance of \textbf{A} vs. \textbf{C}
to
a n\% chance of \textbf{B} vs. \textbf{C}
for \textbf{any} \textbf{C} and \textbf{any} n!
Norms of rational choice

**Continuity:**
If you prefer A to B to C, then for some $n$:
You are indifferent between
having B for sure
and
a n% chance of A vs. C.

Example: **Thai > Mexican > Ukranian**
– How about Mexican, vs. 75/25 Thai/Ukranian?
– How about Mexican, vs. 50/50 Thai/Ukranian?
– How about Mexican, vs. 25/75 Thai/Ukranian?
vNM utility theorem

If your preferences satisfy

1. transitivity
2. irreflexivity
3. completeness
4. independence of irrelevant alts
5. continuity

Then your preferences can be modeled as:

A preferred to B if and only if $EU(A) > EU(B)$
Expected utility (EU)

**P**: a probability function on events $E$

**U**: a function assigning real numbers to fully specified states $s$

**expected utility**: average utility of the true state, if $E$ holds (roughly: your best guess)

$$EU(E) = \sum_{s \in E} U(s) \times P(s \mid E)$$
Want to bet on a coin flip?
Heads: I give you $5,000,000.
Tails: You give me $100.

**Choices:**
– take the bet
– refuse the bet

Suppose utility = money earned.
What is the EU of each choice?
An offer you can’t refuse

Heads: I give you $5m.
Tails: You give me $100.

\[ EU(\text{bet}) = P(\text{flip H}) \times U(\text{bet} & \text{flip H}) + P(\text{flip T}) \times U(\text{bet} & \text{flip T}) \]

\[ = .5 \times 5,000,000 + .5 \times -100 = 2,499,950 \]

\[ EU(\text{don’t bet}) = .5 \times 0 + .5 \times 0 = 0 \]
EU maximization

Bayesian norm of rational choice:

Given some available actions \( a, b, c, d, \ldots \)

a rational Bayesian agent will choose the action with highest expected utility.

\[
EU(a) = \sum_{s \in E} U(s) \times P(s \mid E)
\]
An offer you can’t refuse?

Want to bet on a coin flip?
  Heads: I give you $5,000,000.
  Tails: You give me $100.

Are you rationally obliged to take this bet?
  What if losing $100 would bankrupt you?
  What if you really, really hate losing?
No Bayesian constraints on values

Heads: I give you $5,000,000.
Tails: You give me $100.

Alternatives to ‘utility = earnings’?
1. ‘diminishing marginal utility’
2. drastic effects of small gains/losses
3. valuing future emotional states (e.g., regret)
4. any wacky thing you could dream up
Short vs. long term

How should you choose a restaurant?

1. Always choose the best restaurant you’ve ever been to?
2. Always choose a random restaurant?
3. ‘Explore/exploit’: it depends on how long you will stay in the area!
Value of information

1. Choices affect the world
2. These effects provide information

Example: going to a restaurant tells you something about how good the food is

→ Could this dynamic lead to apparently sub-optimal choices?
Multi-armed bandits

Machine A: 1 jackpot in 3 pulls
Machine B: 50 jackpots in 100 pulls

You’ll be here for a while …

Which do you try next?
Multi-armed bandits

Machine A: 1 jackpot in 3 pulls
Machine B: 50 jackpots in 100 pulls

You have to leave after 1 more pull ...

Which do you try next?
Some more puzzles about choice

Risk aversion
Sunk cost fallacy
Endowment effect
Framing effects
Effects of emotion
Regret, disappointment, …
Endowment effect

Experimenters
1. Chose participants at random to receive mugs or chocolate.
2. Told them they’d been randomly assigned.
3. Asked whether they’d like to trade for the other.

– What do you expect to happen?
Ellsberg

Urn contents:
• 30 red balls
• 60 balls that are either black or yellow

Which gamble do you prefer?
1. $50 if you draw a red ball
2. $50 if you draw a black ball
Ellsberg

Same urn:
- 30 red balls
- 60 balls that are either black or yellow

Which gamble do you prefer?
1. $50 if you draw a red or yellow ball
2. $50 if you draw a black or yellow ball
Next time

Decision and the brain

Brian Knutson

Research Article

Neural Affective Mechanisms Predict Market-Level Microlending

Alexander Genevsky¹ and Brian Knutson¹,²
¹Department of Psychology and ²Stanford Neurosciences Institute, Stanford University