Formal Analysis

Andy Eggers

1/13/2018

Plan for today

- What is formal theory for?

Plan for today

- What is formal theory for?
- Overview of the course & syllabus

Plan for today

- What is formal theory for?
- Overview of the course & syllabus
- Preferences, rationality, utility, expected utility

What is formal theory for?

What are theories?

Theories are things we believe to be true, at least provisionally.

Theories are claims that make sense of regularities/patterns in the world.

What do we do with theories?

- Prediction

What do we do with theories?

- Prediction
  - How would we expect X to affect Y?

What do we do with theories?

- Prediction
  - How would we expect X to affect Y?
- Explanation

What do we do with theories?

- Prediction
  - How would we expect X to affect Y?
- Explanation
  - X and Y are related. Why?

What's wrong with the theory in most dissertations?

- No theory

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions
  - Some assumptions not necessary

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions
  - Some assumptions not necessary
  - Theory doesn't address important features of problem

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions
  - Some assumptions not necessary
  - Theory doesn't address important features of problem
  - Assumptions not justified, too specific

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions
  - Some assumptions not necessary
  - Theory doesn't address important features of problem
  - Assumptions not justified, too specific
- Poorly motivated theory

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions
  - Some assumptions not necessary
  - Theory doesn't address important features of problem
  - Assumptions not justified, too specific
- Poorly motivated theory
  - No puzzle to explain

What's wrong with the theory in most dissertations?

- No theory
  - No reason to expect the predicted pattern, or any pattern
  - No explanation for the finding
- Bad theory
  - Conclusions don't follow from assumptions
  - Some assumptions not necessary
  - Theory doesn't address important features of problem
  - Assumptions not justified, too specific
- Poorly motivated theory
  - No puzzle to explain
  - Prediction obvious or produced by many theories

Theories and theory testing: physics

Hypothetico-deductive ideal: we believe one theory (Newtonian physics); we have another candidate theory (general relativity).

Theories and theory testing: physics

Hypothetico-deductive ideal: we believe one theory (Newtonian physics); we have another candidate theory (general relativity).

We design a **crucial experiment** for which the two theories make different predictions.

Theories and theory testing: physics

Hypothetico-deductive ideal: we believe one theory (Newtonian physics); we have another candidate theory (general relativity).

We design a **crucial experiment** for which the two theories make different predictions.

We run the experiment and discard one of the theories.

Theories and theory testing: social science

Social science reality: The social world is very complicated, so our theories are probabilistic, scope-limited, partial $\rightarrow$ experiments rarely crucial. Theories rarely discarded.

Theories and theory testing: social science

Social science reality: The social world is very complicated, so our theories are probabilistic, scope-limited, partial $\rightarrow$ experiments rarely crucial. Theories rarely discarded.

Often, we want to **clarify** relevant mechanisms, all of which likely play a role; compare their usefulness in prediction/explanation.

Theories and theory testing: social science

Social science reality: The social world is very complicated, so our theories are probabilistic, scope-limited, partial $\rightarrow$ experiments rarely crucial. Theories rarely discarded.

Often, we want to **clarify** relevant mechanisms, all of which likely play a role; compare their usefulness in prediction/explanation.

Physics theory: Elaborating theories to develop crucial experiments, keep only what works best

Theories and theory testing: social science

Social science reality: The social world is very complicated, so our theories are probabilistic, scope-limited, partial $\rightarrow$ experiments rarely crucial. Theories rarely discarded.

Often, we want to **clarify** relevant mechanisms, all of which likely play a role; compare their usefulness in prediction/explanation.

Physics theory: Elaborating theories to develop crucial experiments, keep only what works best

Social science theory: Elaborating theories to improve internal coherence, make judgments about what is useful

Why should theory be formal?

Doesn't have to be.

Why should theory be formal?

Doesn't have to be. But can be useful for

- Development of theory

Why should theory be formal?

Doesn't have to be. But can be useful for

- Development of theory
  - Logical consistency

Why should theory be formal?

Doesn't have to be. But can be useful for

- Development of theory
  - Logical consistency
  - Abstraction $\rightarrow$ connections

Why should theory be formal?

Doesn't have to be. But can be useful for

- Development of theory
  - Logical consistency
  - Abstraction $\rightarrow$ connections
- Communication of theory (depends on the crowd)

Why should theory be formal?

Doesn't have to be. But can be useful for

- Development of theory
  - Logical consistency
  - Abstraction $\rightarrow$ connections
- Communication of theory (depends on the crowd)
  - **Good**: abstraction and simplicity clarify and enlighten

Why should theory be formal?

Doesn't have to be. But can be useful for

- Development of theory
  - Logical consistency
  - Abstraction $\rightarrow$ connections
- Communication of theory (depends on the crowd)
  - **Good**: abstraction and simplicity clarify and enlighten
  - **Bad**: notation and complexity overwhelm and confuse

Types of formal theory

- **Decision theory**: how agents optimize, e.g. partisan voter in event of scandal

Types of formal theory

- **Decision theory**: how agents optimize, e.g. partisan voter in event of scandal
- **Game theory**: how optimizing agents interact, e.g. politician and voter

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)
  - perfectly informed

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)
  - perfectly informed
  - capable of heroic computations

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)
  - perfectly informed
  - capable of heroic computations
- The best model has the most realistic assumptions

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)
  - perfectly informed
  - capable of heroic computations
- The best model has the most realistic assumptions

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)
  - perfectly informed
  - capable of heroic computations
- The best model has the most realistic assumptions

Actually:

- Agents may be optimizing anything, and with constraints on info, processing power, etc

Misconceptions about formal theory

- Formal theory necessarily assumes that agents are
  - selfish (vs. altruistic)
  - materialistic (vs. e.g. idealistic)
  - perfectly informed
  - capable of heroic computations
- The best model has the most realistic assumptions

Actually:

- Agents may be optimizing anything, and with constraints on info, processing power, etc
- "All models are wrong, but some are useful" (George Box)

Pitfalls of formal theory

- Formal theorists like **interesting** theories, but sometimes the best theory is not interesting: e.g. domination.

Pitfalls of formal theory

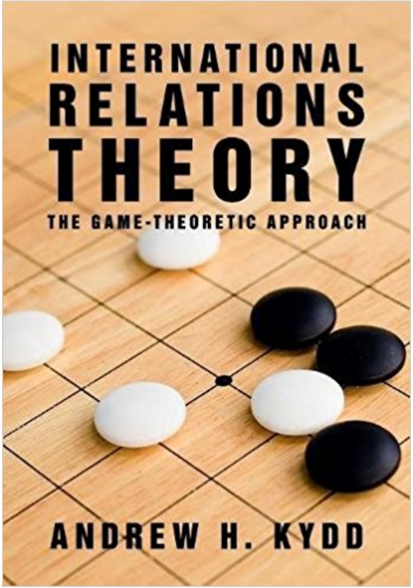
- Formal theorists like **interesting** theories, but sometimes the best theory is not interesting: e.g. domination.
- Modeling assumptions become beliefs about the world.

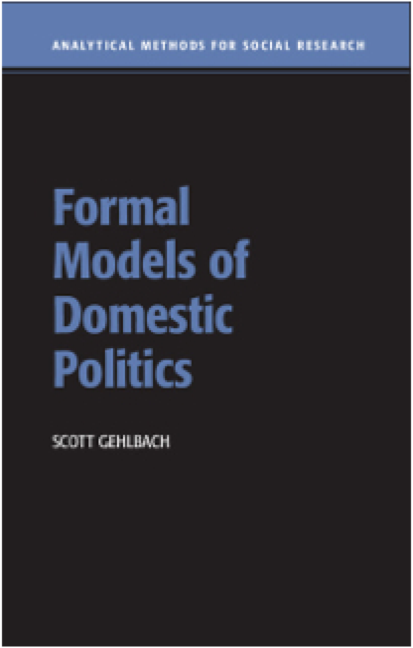
Pitfalls of formal theory

- Formal theorists like **interesting** theories, but sometimes the best theory is not interesting: e.g. domination.
- Modeling assumptions become beliefs about the world.
- They get carried away with technical challenges.

Overview of course

Textbooks


INTERNATIONAL
RELATIONS
THEORY
THE GAME-THEORETIC APPROACH
ANDREW H. KYDD


ANALYTICAL METHODS FOR SOCIAL RESEARCH
Formal
Models of
Domestic
Politics
SCOTT GEHLBACH

Schedule

Wk.Sess	Topic	Reading
1.1	Introduction and utility theory	Kydd 2
1.2	Strategic settings	Kydd 3
2.1	Bargaining	Kydd 4
2.2	Electoral competition under certainty	Gehlbach 1
3.1	Electoral competition under uncertainty	Gehlbach 2
3.2	Application (**Problem set 1 due**, February 1)	
4.1	Power change and war	Kydd 5
4.2	Private information and war	Kydd 6
5.1	Arms competition and war	Kydd 7
5.2	Cooperation theory	Kydd 8
6.1	Application (**Problem set 2 due**, February 20)	
6.2	Special interest politics	Gehlbach 3.1-3.4
7.1	Veto players	Gehlbach 4
7.2	Delegation	Gehlbach 5
8.1	Diplomacy and signaling	Kydd 9
8.2	Application (**Problem set 3 due**, March 8)	

Assessment

Applications and in-class quizzes	10%
Written assignment (due week 9)	20%
Problem sets (due weeks 3, 6, and 8)	30%
Final exam (take-home, due week 0 of TT)	40%

Background

This course will be harder if you've never seen notation like:

$$L = (p_1, \ldots, p_n)$$

$$U(L) \equiv \sum_{i=1}^{n} p_i u(x_i)$$

$$u(x) = x^2 \quad (1)$$

$$u'(x) = 2x \quad (2)$$

For catch-up/review, see Moore & Siegel's *A Mathematics Course for Political & Social Research* (2013) chapters 1, 2, 3, 5, 6, and 8.

Expectations

- You:

Expectations

- You:
  - Do the reading for each session (at least attempt it).

Expectations

- You:
  - Do the reading for each session (at least attempt it).
  - If confused by reading, visit office hours (Tues, Thurs 11-12:30, Nuffield K4) and/or pose questions on Slack

Expectations

- You:
  - Do the reading for each session (at least attempt it).
  - If confused by reading, visit office hours (Tues, Thurs 11-12:30, Nuffield K4) and/or pose questions on Slack
  - If confused by what happens in class, stop me and/or do above

Expectations

- You:
  - Do the reading for each session (at least attempt it).
  - If confused by reading, visit office hours (Tues, Thurs 11-12:30, Nuffield K4) and/or pose questions on Slack
  - If confused by what happens in class, stop me and/or do above
- I:

Expectations

- You:
  - Do the reading for each session (at least attempt it).
  - If confused by reading, visit office hours (Tues, Thurs 11-12:30, Nuffield K4) and/or pose questions on Slack
  - If confused by what happens in class, stop me and/or do above
- I:
  - respond to your questions in office hours and on Slack

Expectations

- You:
  - Do the reading for each session (at least attempt it).
  - If confused by reading, visit office hours (Tues, Thurs 11-12:30, Nuffield K4) and/or pose questions on Slack
  - If confused by what happens in class, stop me and/or do above
- I:
  - respond to your questions in office hours and on Slack
  - design quizzes, activities, mini-lectures, discussions, problem sets that reward your efforts in class and outside of class

Preferences, rationality, utility, expected utility

Preferences

Define set of outcomes $X = \{a, b, c\}$.

- a at least as good as b: $a \succcurlyeq b$.

Preferences

Define set of outcomes $X = \{a, b, c\}$.

- a at least as good as b: $a \succcurlyeq b$.
- a better than b: $a \succ b$.

Preferences

Define set of outcomes $X = \{a, b, c\}$.

- a at least as good as b: $a \succcurlyeq b$.
- a better than b: $a \succ b$.
- a no better or worse than b: $a \sim b$.

Preferences

Define set of outcomes $X = \{a, b, c\}$.

- a at least as good as b: $a \succcurlyeq b$.
- a better than b: $a \succ b$.
- a no better or worse than b: $a \sim b$.

Preferences

Define set of outcomes $X = \{a, b, c\}$.

- a at least as good as b: $a \succcurlyeq b$.
- a better than b: $a \succ b$.
- a no better or worse than b: $a \sim b$.

If preferences are **complete**, $a \succcurlyeq b$ or $b \succcurlyeq a$ (or both) for any pair of alternatives a and b.

If preferences are **transitive**, $a \succcurlyeq b$ and $b \succcurlyeq c$ implies $a \succcurlyeq c$.

Rationality

If behavior is consistent with complete and transitive preferences, it is often called **rational**.

> *The theory of rational choice ... is that a decision-maker chooses the best action according to her preferences, among all the actions available to her. No qualitative restriction is placed on the decision-maker's preferences; her rationality lies in the consistency of her decisions when faced with different sets of available actions, not in the nature of her likes and dislikes. (Osborne 2004, p. 4)*

Quick detour: if and only if, $\iff$, necessary and sufficient

Identical statements:

- **Condition A** is true if and only if **Condition B** is true

Quick detour: if and only if, $\iff$, necessary and sufficient

Identical statements:

- **Condition A** is true if and only if **Condition B** is true
- Condition A $\iff$ Condition B

Quick detour: if and only if, $\iff$, necessary and sufficient

Identical statements:

- **Condition A** is true if and only if **Condition B** is true
- Condition A $\iff$ Condition B
- A is a necessary and sufficient condition for B

Quick detour: if and only if, $\iff$, necessary and sufficient

Identical statements:

- **Condition A** is true if and only if **Condition B** is true
- Condition A $\iff$ Condition B
- A is a necessary and sufficient condition for B
- A is true whenever B is true and vice versa

Ordinal utility

Consider representing an agent's preferences over X by assigning a score to each outcome.

Ordinal utility

Consider representing an agent's preferences over X by assigning a score to each outcome.

Definition 2.2 (Kydd) A function $u : X \to \mathbb{R}$ is a utility function representing the preferences $\succcurlyeq$ if (and only if), for all $x_i, x_j \in X$, $u(x_i) \geq u(x_j) \iff x_i \succcurlyeq x_j$.

Ordinal utility

Consider representing an agent's preferences over X by assigning a score to each outcome.

Definition 2.2 (Kydd) A function $u : X \to \mathbb{R}$ is a utility function representing the preferences $\succcurlyeq$ if (and only if), for all $x_i, x_j \in X$, $u(x_i) \geq u(x_j) \iff x_i \succcurlyeq x_j$.

- Can a utility function be found for any set of preferences?

Ordinal utility

Consider representing an agent's preferences over X by assigning a score to each outcome.

Definition 2.2 (Kydd) A function $u : X \to \mathbb{R}$ is a utility function representing the preferences $\succcurlyeq$ if (and only if), for all $x_i, x_j \in X$, $u(x_i) \geq u(x_j) \iff x_i \succcurlyeq x_j$.

- Can a utility function be found for any set of preferences?
- Find a utility function for these preferences: $a \succ b$, $b \sim c$, $c \succ d$

Choice under uncertainty

Sometimes we are not choosing among outcomes $\{a, b, c\}$, but rather among actions $\{1, 2, 3\}$ that probabilistically lead to one of those outcomes

Example: Voting.

Each action leads to a *lottery* over alternatives.

Definition 2.3 (Kydd) A lottery associated with a finite set of outcomes, X, with number of elements equal to $|X| = n$, is a vector $L = (p_1, \ldots, p_n)$, where $p_i \in [0, 1]$ is interpreted as the probability that outcome i occurs, so that $\sum_i p_i = 1$.

Expected utility

The expected utility of a lottery is the *probability-weighted average utility* of the outcomes.

Expected utility

The expected utility of a lottery is the *probability-weighted average utility* of the outcomes.

Definition 2.4 (Kydd) The expected utility of a lottery L based on finite outcome set X is defined as the expected value of the utilities of the outcomes

$$EU(L) \equiv \sum_{i=1}^{n} p_i u(x_i)$$

Expected utility

The expected utility of a lottery is the *probability-weighted average utility* of the outcomes.

Definition 2.4 (Kydd) The expected utility of a lottery L based on finite outcome set X is defined as the expected value of the utilities of the outcomes

$$EU(L) \equiv \sum_{i=1}^{n} p_i u(x_i)$$

What is the expected utility of $L = \{.2, .3, .5\}$, given $u(x) = \{3, 2, 1\}$?

Preferences over lotteries

Suppose a set of lotteries $\mathcal{L}$. Agents can have preferences over these lotteries: e.g. $L_1 \succ L_2$.

Preferences over lotteries

Suppose a set of lotteries $\mathcal{L}$. Agents can have preferences over these lotteries: e.g. $L_1 \succ L_2$.

If these preferences are complete and transitive, we can find a utility function $v : \mathcal{L} \to \mathbb{R}$ to represent those preferences.

Preferences over lotteries

Suppose a set of lotteries $\mathcal{L}$. Agents can have preferences over these lotteries: e.g. $L_1 \succ L_2$.

If these preferences are complete and transitive, we can find a utility function $v : \mathcal{L} \to \mathbb{R}$ to represent those preferences.

Wouldn't it be great if there were a utility function over **outcomes** such that preferences over **lotteries** were given by the associated **expected utilities**?

Preferences over lotteries

Suppose a set of lotteries $\mathcal{L}$. Agents can have preferences over these lotteries: e.g. $L_1 \succ L_2$.

If these preferences are complete and transitive, we can find a utility function $v : \mathcal{L} \to \mathbb{R}$ to represent those preferences.

Wouldn't it be great if there were a utility function over **outcomes** such that preferences over **lotteries** were given by the associated **expected utilities**?

i.e. $EU(L_1) \geq EU(L_2) \iff L_1 \succcurlyeq L_2$ for all $L_1, L_2 \in \mathcal{L}$.

Preferences over lotteries

Suppose a set of lotteries $\mathcal{L}$. Agents can have preferences over these lotteries: e.g. $L_1 \succ L_2$.

If these preferences are complete and transitive, we can find a utility function $v : \mathcal{L} \to \mathbb{R}$ to represent those preferences.

Wouldn't it be great if there were a utility function over **outcomes** such that preferences over **lotteries** were given by the associated **expected utilities**?

i.e. $EU(L_1) \geq EU(L_2) \iff L_1 \succcurlyeq L_2$ for all $L_1, L_2 \in \mathcal{L}$.

This can't be true for all imaginable preferences.

Preferences over lotteries

Suppose a set of lotteries $\mathcal{L}$. Agents can have preferences over these lotteries: e.g. $L_1 \succ L_2$.

If these preferences are complete and transitive, we can find a utility function $v : \mathcal{L} \to \mathbb{R}$ to represent those preferences.

Wouldn't it be great if there were a utility function over **outcomes** such that preferences over **lotteries** were given by the associated **expected utilities**?

i.e. $EU(L_1) \geq EU(L_2) \iff L_1 \succcurlyeq L_2$ for all $L_1, L_2 \in \mathcal{L}$.

This can't be true for all imaginable preferences.

But John von Neumann and Oscar Morgenstern (1947) proved that there is such a utility function if and only if preferences over lotteries are complete, transitive, continuous, and independent (**Theorem 2.2 in Kydd**).

Cardinal utilities, i.e. von Neumann Morgenstern (VNM) utilities

Morgenstern and Von Neumann, 1946

To show: if $u(x)$ is a VNM (cardinal) utility function (i.e. expected utility of lotteries tracks preferences over lotteries), then so is $a + bu(x)$, where $b > 0$.

What does this mean about cardinal utilities, in plain English?

Proof

To show: if $u(x)$ is a VNM (cardinal) utility function, then so is $a + bu(x)$, where $b > 0$.

Proof Call the expected utility of a lottery L under the original utility function $U(L)$, and call the expected utility a lottery L under the transformed utility function $V(L)$. We need to show that $U(L) \geq U(L') \iff V(L) \geq V(L')$ for all L, L'.

First we show $V(L) = a + bU(L)$. Recall $U(L) \equiv \sum_{i=1}^{n} p_i u(x_i)$. Observe that

$$V(L) \equiv \sum_{i=1}^{n} p_i \big(a + bu(x_i)\big) \tag{3}$$

$$= a \sum_{i=1}^{n} p_i + b \sum_{i=1}^{n} p_i u(x_i) \tag{4}$$

$$= a + bU(L). \tag{5}$$

Now, suppose that $U(L) \geq U(L')$ for some L, L'. Then $bU(L) \geq bU(L')$, assuming $b > 0$. And $a + bU(L) \geq a + bU(L')$ for any a. This implies that $U(L) \geq U(L') \iff V(L) \geq V(L')$. QED.

Cardinal utilities: relative values matter, but not scale or location

With **cardinal** utilities,

- the relative values matter: if $\{3, 2, 1\}$ works as cardinal utility, $\{3, 2, -1000\}$ does not.
- the scale does not matter: if $\{3, 2, 1\}$ works as cardinal utility, $\{300, 200, 100\}$ and $\{5, 4, 3\}$ and $\{2, 1, 0\}$ do too.

Cardinal utilities: relative values matter, but not scale or location

With **cardinal** utilities,

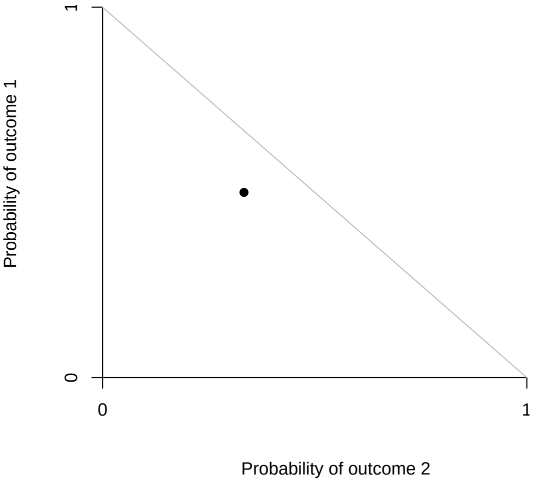
- the relative values matter: if $\{3, 2, 1\}$ works as cardinal utility, $\{3, 2, -1000\}$ does not.
- the scale does not matter: if $\{3, 2, 1\}$ works as cardinal utility, $\{300, 200, 100\}$ and $\{5, 4, 3\}$ and $\{2, 1, 0\}$ do too.

So we can e.g. **normalize** to 0-1 scale.

Intuitive sense of VNM's theorem (1)

Suppose there are only three outcomes.

Lotteries can be depicted as points on the **simplex**:



Intuitive sense of VNM's theorem (2)

Normalize utilities such that $u_1 = 1$ and $u_3 = 0$.

Intuitive sense of VNM's theorem (2)

Normalize utilities such that $u_1 = 1$ and $u_3 = 0$.

Then $EU(L) = p_1 + p_2 u_2$.

Intuitive sense of VNM's theorem (2)

Normalize utilities such that $u_1 = 1$ and $u_3 = 0$.

Then $EU(L) = p_1 + p_2 u_2$.

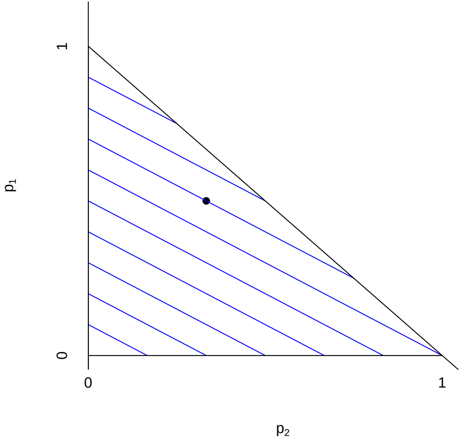
Rearranging, $p_1 = U(L) - p_2 u_2$.

Intuitive sense of VNM's theorem (2)

Normalize utilities such that $u_1 = 1$ and $u_3 = 0$.

Then $EU(L) = p_1 + p_2 u_2$.

Rearranging, $p_1 = U(L) - p_2 u_2$. We can plot this line for different values of $U(L)$:



Intuitive sense of VNM's theorem (3)

Definition: Preferences over lotteries are *independent* if

$$L \succcurlyeq L^{'} \iff \alpha L + (1-\alpha)L^{''} \succcurlyeq \alpha L^{'} + (1-\alpha)L^{''}$$

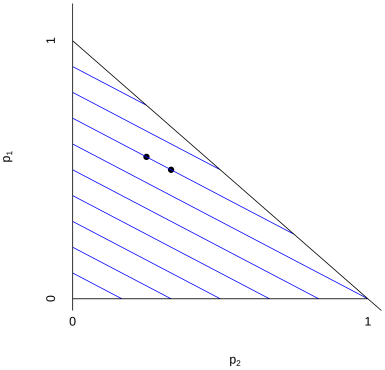
Kydd p. 15: "If $L_1 \succ L_2$, then adding an equal chance of obtaining L_3 to both sides does not alter the preference."

Intuitive sense of VNM's theorem (4)

Independence implies that

$$L \sim L' \iff \alpha L + (1-\alpha)L'' \sim \alpha L' + (1-\alpha)L''$$

And this implies that the **indifference curves** follow the same pattern as the expected utilities. Therefore preferences that satisfy the axioms can be represented by expected utility.

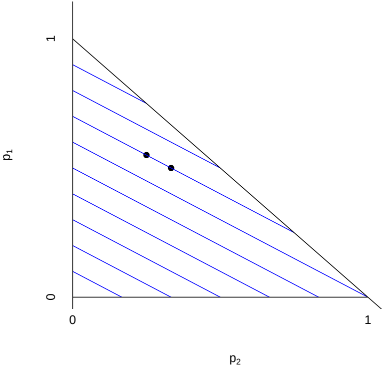


Intuitive sense of VNM's theorem (4)

Independence implies that

$$L \sim L^{'} \iff \alpha L + (1-\alpha)L^{''} \sim \alpha L^{'} + (1-\alpha)L^{''}$$

And this implies that the **indifference curves** follow the same pattern as the expected utilities. Therefore preferences that satisfy the axioms can be represented by expected utility.



Google "Jonathan Levin choice under uncertainty" for more rigorous version.

Risk preferences