probability and counterfactuals

fabrizio cariani and paolo santorio

northwestern + maryland

august 6, 2019

Lecture 2

a recap of yesterday: semantics

- counterfactuals are modal sentences
- standardly modeled in possible world semantics
- standard theory: A > C is true at w iff C is true at all the closest relevant worlds in which A is true
- Formally:

$$[\![\text{A}> \text{C}]\!]^w = 1 \text{ iff } \forall v \in max_{\leq_w}(f(w) \cap [\![\text{A}]\!]^{v,f,\leq}), [\![\text{C}]\!]^{v,f,\leq} = 1$$

a recap of yesterday: probability

- probability is involved in the analysis of subjective credence and of objective chance
- agents have numerically representable states of subjective credence
- requirements of rationality on agent a:
  - a's credence functions is a probability function p
  - a's conditional credences are the conditional probabilities of p
  - a updates by conditionalization
- chance is an objective, agent-insensitive concept of probability

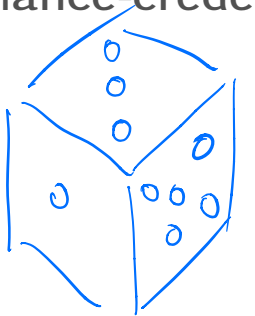
} standard Bayesian

today's goals & outline

1. one left-over topic from yesterday: the relation between chance and credence
2. introduce Stalnaker's Thesis; a notorious constraint on the probability of indicative conditionals
3. introduce, modify and critique, Skyrms's Thesis; a less famous constraint on the probability of counterfactuals
4. modify Skyrms's Thesis and consider preliminary objections

chance-credence principles

chance-credence principles: a puzzle



it's a fair die!

c(it will land on 3) $\doteq .8$

chance-credence principles: introduction

- are there principles of rationality in addition to probabilism?
- one candidate: chance-credence principles
- informally, principles that regulate the coherence of
  - credences about chancy-events
  - credences about the chances of those events

example 1

it is incoherent (and hence irrational) to

- believe (=have credence 1) that the die is fair while
- have credence .8 that you will roll a six

example 2

it is incoherent (and hence irrational) to

- believe (=have credence 1) that there is a .95 objective chance of Steph Curry will make his next free throw

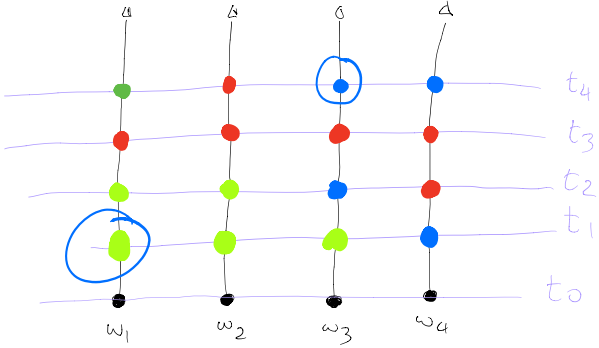
  while

- have credence .3 that he will

auxiliary concepts 1

history propositions

H_t^w is a sentence s.t. $[\![H_t^w]\!] = \{v \mid v \text{ is a duplicate of } w \text{ up to } t\}$



IN THIS EXAMPLE:

$H^{w_1}_{t_1} = \{w_1, w_2, w_3\}$

$H^{w_1}_{t_2} = \{w_1, w_2\}$

$H^{w_1}_{t_4} = \{w_1\}$

auxiliary concept 2

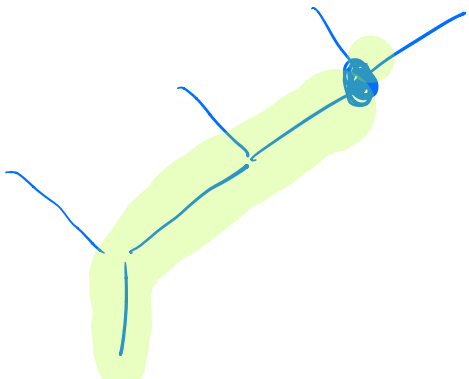
theory-of-chance propositions

T^w is a sentence that states all of the chances of every chancy event in w at any time in w

auxiliary concepts 3

chance-grounding claims (CGC)

conjunctions of the form $H_t^w \ \& \ T^w$ are very powerful. Call such propositions chance-grounding claims (CGC's)



the principal principle (PP)

- initially proposed by Lewis (1986), who gave two different formulations of it
- Meacham (2010) argues persuasively that one of them is superior

formulating PP

- let *ic* be an agent's *initial* credence (remember *superbabies*).
- then say that the Principal Principle is the constraint that for all w and t,

$$ic(\mathrm{A} \mid H_t^w \ \& \ T^w) = ch_{w,t}(\mathrm{A})$$

- informally, the initial credence of an agent conditional on a CGC (at w, t) equals the chance of A at t

Q&A about $ic(\mathrm{A} \mid H_t^w \ \& \ T^w) = ch_{w,t}(\mathrm{A})$

Q: does PP entail that if the chance of A at t in w is .3, then a rational agent's credence in A at t in w is also .3?

A: no. PP only directly constrains the agent's *initial* credence. A rational agent located at t in w might not be at the beginning of her epistemic life.

Q&A about $ic(\mathrm{A} \mid H_t^w \ \& \ T^w) = ch_{w,t}(\mathrm{A})$

Q: Formulations of PP usually involve a notion of "admissible evidence"? Why does your version not involve one?

A: The formulations that need that restriction look like this:

$$ic(\mathrm{A} \mid \text{"the chance of A is } r\text{"} \ \& \ \mathrm{E}) = r$$

This is more general than our formulation because E could be anything. The admissibility clause helps limit its scope. We don't need it.

Q&A about $ic(\mathrm{A} \mid H_t^w \ \& \ T^w) = ch_{w,t}(\mathrm{A})$

Q: What if the agent's evidence is weaker than a CGC?

A: Then our formulation *indirectly* constrains the agent's credence.

- Suppose you learn that a certain coin is fair.
- That is much weaker than learning a whole CGC.
- However, it's equivalent to learning a disjunction of CGC's.
- ... but even so Bayesianism constrains your credence
- advance to the next slide to see how!

- IF you know the values of $ic(A \mid C_i)$ for each $i \in \{1, ..., n\}$ and for pairwise incompatible C_1,...,C_n
- THEN $ic(A \mid C_1 \vee ... \vee C_n)$ will be a weighted average of those values.

$$ic(A \mid C_1 \vee ... \vee C_n) = \sum_{1 \leq i \leq n} ic(A \mid C_i) \cdot \frac{ic(C_i)}{ic(C_1 \vee ... \vee C_n)}$$

example

- at t_1, you are a superbaby; you know nothing, but hypothesize a lot and your credence c is rational.
- at t_2, you learn that a coin that is to be flipped is fair (and nothing else).
- by conditionalization, $c_2(\cdot) = c_1(\cdot \mid \text{"the coin is fair"})$.
- the PP did not directly set $c_1(\cdot \mid \text{"the coin is fair"})$ but...
- ... it did set $c_1(\cdot \mid H_t^w \ \& \ T^w)$.
- identify all the CGC's that agree that the coin is fair.
- set $c_1(\cdot \mid \text{"the coin is fair"})$ to the weighted average of the CGC's.

stalnaker's thesis

Stalnaker's Thesis (aka *the Thesis*)

For all c that model rational credence, and for all A, B s.t. $Pr(A) > 0$:

$$c(A \rightarrow C) = c(C \mid A)$$

Two interpretations:

- normative
- descriptive

Why believe the thesis? Intuitive arguments.

Suppose that Maria might have tossed a fair die, and assess the probability of (1):

(1) If Maria tossed the die, it landed on 1 or 2.

The natural answer is '1/3', which of course is also the value of the corresponding conditional probability.

Lots of experiments support this.

Why believe the thesis? A derivation.

Assume:

- probabilistic centering:

$$c((A \rightarrow B) \ \& \ A) = c(A \ \& \ B)$$

  Indeed, $(A \rightarrow B) \ \& \ A$ is equivalent to $A \ \& \ B$ depending only on very basic assumptions of conditional logic (Strong + Weak Centering).

- independence: for all rational c and all A, B s.t. $c(A) > 0$,

$$c(A \rightarrow B) = c(A \rightarrow B \,|\, A)$$

Assume that c is a rational credence function:

i. $c(A \rightarrow B) =$

ii. $c(A \rightarrow B \mid A) =$ (via Independence)

iii. $\dfrac{c(A \rightarrow B \wedge A)}{c(A)} =$ (def of conditional probability)

iv. $\dfrac{c(A \wedge B)}{c(A)} =$ (Probabilistic Centering)

v. $c(B \mid A)$ (def of conditional probability)

more on the thesis tomorrow

despite these arguments, the thesis is notoriously hard to vindicate

the probabilities of counterfactuals

do counterfactuals have probabilities?

- in some setups, there are strong and natural ascriptions of probabilities to counterfactuals.
- suppose that I have a coin in my pocket with 80% / 20% bias towards heads; instead of flipping it, I decide to melt it.
- in this scenario (2a) is naturally judged more likely than (2b)

(2) a. If I had flipped the coin, it would have landed heads

b. If I had flipped the coin, it would have landed tails

do counterfactuals have probabilities?

P1. counterfactuals express propositions

P2. every proposition has a probability

C. the propositions expressed by counterfactuals have probability

note: P1 and P2 are widely though not universally believed.

two questions

Q1. can we recycle the thesis to account for our judgments about the probabilities of counterfactuals?

Q2. if not, are there analogues of the thesis?

That is: are there principles that constrain an agent's credence in counterfactuals as a function of other features of that agent's credal state?

recycling the thesis?

that is, couldn't we just go with the following?

$$c(A \boxdot\!\!\rightarrow C) = c(C \mid A)$$

reconsider:

(3) If Oswald didn't kill Kennedy, no one else did.

(4) If Oswald hadn't killed Kennedy, no one else would have.

F | low
T | high

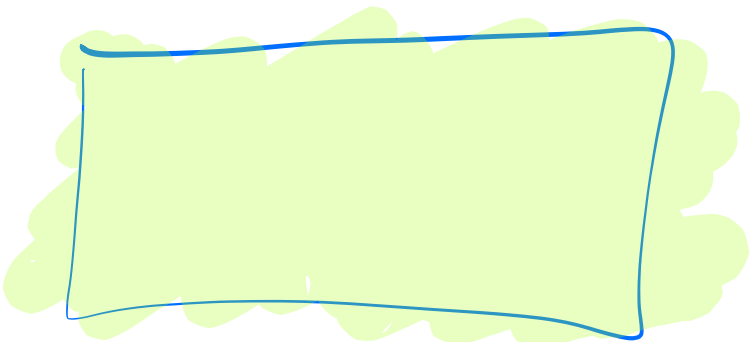
these plausibly differ in truth-value and in credence
the conditional probability $c(\textit{no else did} \mid \textit{not Oswald})$ is low
so it can't be right for (4)

idea one: counterfactual antecedents?

maybe conditioning on a counterfactual antecedent is special

- $c(\cdot \mid \text{"Oswald didn't kill Kennedy"}) \neq c(\cdot \mid \text{"Oswald hadn't killed Kennedy"})$
- that could account for the difference
- however, it doesn't work to model the case correctly...

if we are certain that Kennedy was killed, there is no proposition A that you can learn such that, conditional on A it is highly likely that no one else would have killed Kennedy.



idea two: counterfactual supposition?

maybe "conditional probability" is not the right concept?
what if we distinguish two kinds of suppositions?

- *indicative supposition*: modeled by standard conditional probability $c(\cdot \mid \cdot)$
- *counterfactual supposition*: modeled by "counterfactaul supposing"

we will say a bit more about this tomorrow

idea three: Skyrms' suggestion

My suggestion is that [...] the probabilities involved [in the evaluation of a counterfactual] are the prior propensities rather than the prior epistemic probabilities. [...] If we do not know for certain the values of the prior propensities, we may have to do with a weighted average—the expected prior propensities. The weights in this average will be epistemic probabilities and we should use the best ones available — for this job – the posterior epistemic probabilities.

Formalizing the Vintage Skyrms' Thesis

I will use PR for epistemic probabilities and pr for propensities. I will superscript i. I will superscript i and f for prior (initial) and posterior (final) respectively. Let the double arrow $\Rightarrow$ symbolise the subjunctive conditional, and BAV be 'Basic Assertability Value'. The the theory that I am suggesting can be succinctly expressed thusly:

$$BAV(A \Rightarrow B) = \sum_j PR^f[pr^i_j] \cdot pr^i_j(B \mid A)$$

modding Skyrms 1

$$BAV(\mathrm{A} > \mathrm{B}) = \sum_j PR^f[pr_j^i] \cdot pr_j^i(\mathrm{B} \mid \mathrm{A})$$

modding Skyrms 2

$$c(\mathrm{A} > \mathrm{B}) = \sum_j PR^f[pr^i_j] \cdot pr^i_j(\mathrm{B} \mid \mathrm{A})$$

modding Skyrms 3

$$c(A > B) = \sum_j c[pr_j^i] \cdot pr_j^i(B \mid A)$$

modding Skyrms 4

Let **CH** be the class of (relevant) prior chance functions.
For $ch \in \mathbf{CH}$, let $\chi(ch)$ be the proposition that ch models actual chances (at the salient time)

$$c(\mathrm{A} > \mathrm{B}) = \sum_{ch \in \mathbf{CH}} c[\chi(ch)] \cdot ch(\mathrm{B} \mid \mathrm{A})$$

critiquing modded Skyrms 1

look at that equation again:

$$c(\mathsf{A} > \mathsf{B}) = \sum_{ch \in \mathbf{CH}} c[\chi(ch)] \cdot ch(\mathbf{B} \mid \mathbf{A})$$

- what are the "relevant prior chance functions?"
- and what is the "salient time"?

critiquing modded Skyrms 2

$$c(\mathsf{A} > \mathsf{B}) = \sum_{ch \in \mathbf{CH}} c[\chi(ch)] \cdot ch(\mathbf{B} \mid \mathbf{A})$$

- this principle is reminiscent of the principal principle but ...
- ... it's not just constraining the *initial* credence of the agent
- instead, it is constraining *every credal stage*

critiquing modded Skyrms 3

$$c(A > B) = \sum_{ch \in \mathbf{CH}} c[\chi(ch)] \cdot ch(B \mid A)$$

- suppose I become certain that A and B are both true
- then I should be certain that A > B is true
- but that seems consistent with your giving some credence to some chance functions that do not have $ch(B \mid A)$

towards a counterfactual principal principle

- for these reasons, we might want to try to state a constraint on counterfactual credence in the style of the principal principle (in particular constraining initial credence and not always making "chance information" overriding).
- *note:* we borrow the idea of a counterfactual principal principle from Schulz's recent book *Counterfactuals and Probability* although our development of this principle is different from Schulz's.

the counterfactual principal principle (CPP)

$$ic(\mathsf{A} > \mathsf{C} \mid H_t^w \ \& \ T^w) = ch_{w,t}(\mathsf{C} \mid \mathsf{A})$$

sample application of CPP 1

if Shiny had been flipped on Monday, it would have landed heads

sample application of CPP 2

if Shiny had been flipped on Tuesday, it would have landed heads

sample application of CPP 3

if Oswald hadn't killed Kennedy, no one else would have done it