Logical and Probabilistic Models of Belief Change

Eric Pacuit

Department of Philosophy
University of Maryland, College Park
pacuit.org

August 9, 2017

Plan

- ✓ Introduction to belief revision, AGM
- ✓ Possible worlds models, Bayesian models
- Day 3 Updating probabilities,The value of learning, Lottery Paradox, Preface Paradox, Review Paradox
- Day 4 Lottery Paradox, Preface Paradox, Review Paradox, Iterated belief revision, Context shifts, Becoming aware
- Day 5 Interactive epistemology (Agreement Theorems), Convergence Theorems

$$K_0 \qquad \Longrightarrow \qquad K_t$$

Learn that φ
Suppose that φ

$$K_0 \quad \Longrightarrow \quad K_t = K_0 * \varphi$$

Learn that φ

Suppose that φ

$$K_0 \quad \Longrightarrow \quad K_t = K_0 * \varphi$$

$$p_0 \quad \Longrightarrow \quad p_t$$

$$K_0 \overset{\substack{\text{Learn that } \varphi \\ \text{Suppose that } \varphi}}{\Longrightarrow} K_t = K_0 * \varphi$$

$$p_0 \overset{\text{Learning experience}}{\Longrightarrow} p_t = ???$$

Learn that φ

Suppose that φ

$$K_0 \Longrightarrow K_t = K_0 * \varphi$$

$$\updownarrow \qquad\qquad \updownarrow$$

Learning experience

$$p_0 \Longrightarrow p_t = ???$$

Conditioning

$$p_0 \quad \overset{\text{Learn that } E}{\Longrightarrow} \quad p_t(\cdot) = p_0(\cdot \mid E)$$

Conditional Probability

The probability of E *given* F, dented $p(E|F)$, is defined to be

$$p(E|F) = \frac{p(E \cap F)}{p(F)}.$$

provided $P(F) > 0$.

Setting $p_t(\cdot) = p_0(\cdot \mid E)$ is demonstrably the correct thing to do just in case, for all propositions $H \in \Sigma$, both:

1. Certainty: $p_t(E) = 1$
2. Rigidity: $p_t(H \mid E) = p_0(H \mid E)$

Conditional Probability

The probability of E *given* F, dented $p(E|F)$, is defined to be

$$p(E|F) = \frac{p(E \cap F)}{p(F)}.$$

Conditional Probability

The probability of E *given* F, dented $p(E|F)$, is defined to be

$$p(E|F) = \frac{p(E \cap F)}{p(F)}.$$

Bayes Theorem: $p(E|F) = p(F|E)\frac{p(E)}{p(F)}$

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(Dice \mid Twelve)$ with $p(Roulette \mid Twelve)$.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(Dice \mid Twelve)$ with $p(Roulette \mid Twelve)$. Based on our background knowledge of gambling we have $p(Twelve \mid Dice) = 1/36$ and $p(Twelve \mid Roulette) = 1/38$.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(\mathit{Dice} \mid \mathit{Twelve})$ with $p(\mathit{Roulette} \mid \mathit{Twelve})$. Based on our background knowledge of gambling we have $p(\mathit{Twelve} \mid \mathit{Dice}) = 1/36$ and $p(\mathit{Twelve} \mid \mathit{Roulette}) = 1/38$. Based on our observations about the casino, we can judge the prior probabilities $p(\mathit{Dice})$ and $p(\mathit{Roulette})$.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(Dice \mid Twelve)$ with $p(Roulette \mid Twelve)$. Based on our background knowledge of gambling we have $p(Twelve \mid Dice) = 1/36$ and $p(Twelve \mid Roulette) = 1/38$. Based on our observations about the casino, we can judge the prior probabilities $p(Dice)$ and $p(Roulette)$. But this is now enough to *calculate* the required probabilities.

People are often not aware of all that they have learnt or they fail to adequately represent it, and it is only the failure of the Rigidity condition that alerts us to this.

Three Prisoner's Problem

Three prisoners A, B and C have been tried for murder and their verdicts will told to them tomorrow morning. They know only that one of them will be declared guilty and will be executed while the others will be set free. The identity of the condemned prisoner is revealed to the very reliable prison guard, but not to the prisoners themselves. Prisoner A asks the guard "Please give this letter to one of my friends — to the one who is to be released. We both know that at least one of them will be released".

Three Prisoner's Problem

An hour later, A asks the guard "Can you tell me which of my friends you gave the letter to? It should give me no clue regarding my own status because, regardless of my fate, each of my friends had an equal chance of receiving my letter." The guard told him that B received his letter.

Prisoner A then concluded that the probability that he will be released is $1/2$ (since the only people without a verdict are A and C).

Three Prisoner's Problem

But, A thinks to himself:

Three Prisoner's Problem

But, A thinks to himself:

> *Before I talked to the guard my chance of being executed was 1 in 3. Now that he told me B has been released, only C and I remain, so my chances of being executed have gone from 33.33% to 50%. What happened? I made certain not to ask for any information relevant to my own fate...*

Three Prisoner's Problem

But, A thinks to himself:

> *Before I talked to the guard my chance of being executed was 1 in 3. Now that he told me B has been released, only C and I remain, so my chances of being executed have gone from 33.33% to 50%. What happened? I made certain not to ask for any information relevant to my own fate...*

Explain what is wrong with A's reasoning.

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) =$$

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) = p(I_B \mid G_A)\frac{p(G_A)}{p(I_B)} =$$

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) = p(I_B \mid G_A)\frac{p(G_A)}{p(I_B)} = 1 \cdot \frac{1/3}{2/3} = 1/2$$

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

Given that $p(I'_B \mid G_A)$ is $1/2$ (given that A is guilty, there is a 50-50 chance that the guard could have given the letter to B or C).

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

Given that $p(I'_B \mid G_A)$ is $1/2$ (given that A is guilty, there is a 50-50 chance that the guard could have given the letter to B or C). This gives us the following correct calculation:

$$p(G_A \mid I'_B) = p(I'_B \mid G_A)\frac{p(G_A)}{p(I'_B)} = 1/2 \cdot \frac{1/3}{1/2} = 1/3$$

Setting $p_t(\cdot) = p_0(\cdot \mid E)$ is demonstrably the correct thing to do just in case, for all propositions $H \in \Sigma$, both:

1. Certainty: $p_t(E) = 1$
2. Rigidity: $p_t(H \mid E) = p_0(H \mid E)$

Observation by candlelight

An agent inspects a piece of cloth by candlelight, and gets the impression that it is green (G), although he concedes that it might be blue (B) or even (but very improbably) violet (V).

Observation by candlelight

An agent inspects a piece of cloth by candlelight, and gets the impression that it is green (G), although he concedes that it might be blue (B) or even (but very improbably) violet (V).

$$p_0(G) = p_0(B) = 0.3, \ p_0(V) = 0.4$$

$$\Downarrow$$

$$p_t(G) = 0.7, \ p_t(B) = 0.25, \ p_t(V) = .05$$

Observation by candlelight

An agent inspects a piece of cloth by candlelight, and gets the impression that it is green (G), although he concedes that it might be blue (B) or even (but very improbably) violet (V).

$$p_0(G) = p_0(B) = 0.3,\ p_0(V) = 0.4$$

$$\Downarrow$$

$$p_t(G) = 0.7,\ p_t(B) = 0.25,\ p_t(V) = .05$$

Is there a proposition E such that $p_t(\cdot) = p_0(\cdot \mid E)$?

Jeffrey Conditionalization

When an observation bears directly on the probabilities over a partition $\{E_i\}$, changing them from $p(E_i)$ to $q(E_i)$, the new probability for any proposition H should be

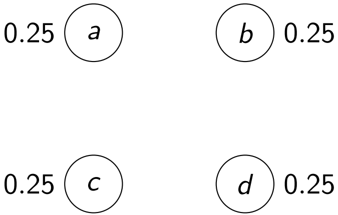
$$q(H) = \sum_i p(H \mid E_i) q(E_i)$$

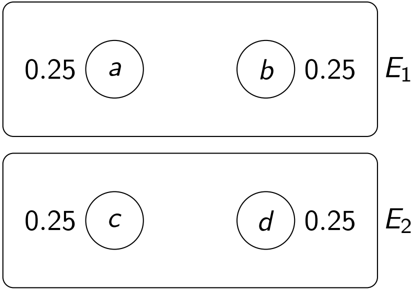
Jeffrey Conditionalization

When an observation bears directly on the probabilities over a partition $\{E_i\}$, changing them from $p(E_i)$ to $q(E_i)$, the new probability for any proposition H should be

$$q(H) = \sum_i p(H \mid E_i)q(E_i)$$

Fact: If q is obtained from p by Jeffrey Conditioning on the partition $\{E, \overline{E}\}$ with $q(E) = 1$, then $q(\cdot) = p(\cdot \mid E)$.

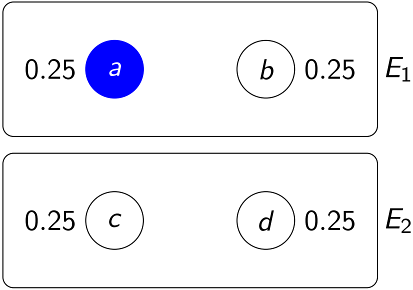

0.25
a
b
0.25
0.25
c
d
0.25



The probability that the guilty party is left-handed is 0.8

$E_1 = \{a, b\}$, $E_2 = \{c, d\}$

$$p(E_1) = 0.8 \qquad p(E_2) = 0.2$$

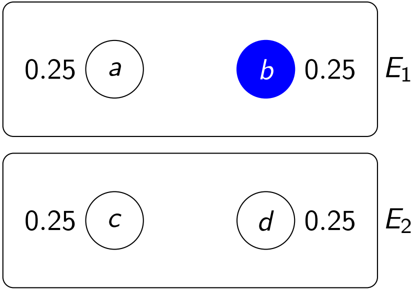


The probability that the guilty party is left-handed is 0.8

$$E_1 = \{a, b\},\ E_2 = \{c, d\}$$

$$p(E_1) = 0.8 \qquad p(E_2) = 0.2$$

$$p(a) = p_0(\{a\} \mid E_1) * p(E_1) + p_0(\{a\} \mid E_2) * p(E_2) = 0.5 * 0.8 + 0 = 0.4$$

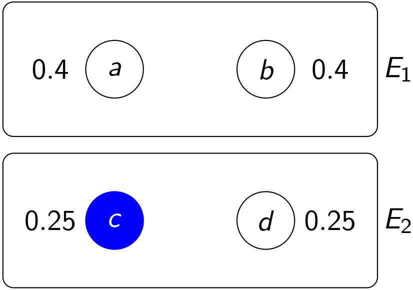


The probability that the guilty party is left-handed is 0.8

$E_1 = \{a, b\}$, $E_2 = \{c, d\}$

$$p(E_1) = 0.8 \qquad p(E_2) = 0.2$$

$$p(b) = p_0(\{b\} \mid E_1) * p(E_1) + p_0(\{b\} \mid E_2) * p(E_2) = 0 + 0.5 * 0.8 = 0.4$$

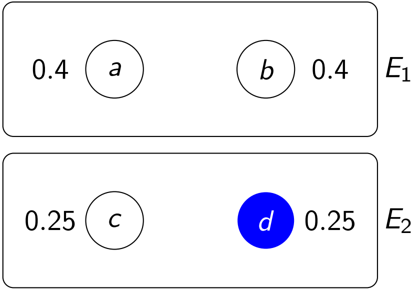


The probability that the guilty party is left-handed is 0.8

$E_1 = \{a, b\},\ E_2 = \{c, d\}$

$p(E_1) = 0.8 \qquad p(E_2) = 0.2$

$p(c) = p_0(\{c\} \mid E_1) * p(E_1) + p_0(\{c\} \mid E_2) * p(E_2) = 0 + 0.5 * 0.2 = 0.1$



The probability that the guilty party is left-handed is 0.8

$E_1 = \{a, b\},\ E_2 = \{c, d\}$

$$p(E_1) = 0.8 \qquad p(E_2) = 0.2$$

$$p(d) = p_0(\{d\} \mid E_1) * p(E_1) + p_0(\{d\} \mid E_2) * p(E_2) = 0 + 0.5 * 0.2 = 0.1$$

P. Diaconis and S. Zabell. *Updating Subjective Probability*. Journal of the American Statistical Association, Vol. 77, No. 380., pp. 822-830 (1982).

Fact. Jeffrey conditioning is not commutative.

Commutativity on Experiences Any rule for updating degrees of belief on experiences should be such that the result of updating credences on one experience and then another should be the same as the result of updating on the same two experiences in reverse order.

Holism For any experience and any proposition, there is a "defeater" proposition, such that your degree of belief in the first proposition, upon having the experience, should depend on your degree of belief in the defeater proposition.

J. Weisberg. *Commutativity or Holism? A Dilemma for Conditionalizers*. British Journal of the Philosophy of Science, 60(4), pp. 793-812, 2009.

M. Lange. *Is Jeffrey Conditionalization Defective in Virtue of Being NonCommutative? Remarks on the Sameness of Sensory Experience*. Synthese 123: 393-403, 2000.

C. Wagner. *Probability kinematics and commutativity*. Philosophy of Science 69, 266-278, 2002.

Conditioning

- If q comes from p by conditioning on E, then $q(E) = 1$

Conditioning

- If q comes from p by conditioning on E, then $q(E) = 1$
- If $p(E_1) = 1$ then $p(E_1 \mid E_2) = 1$ for any E_2 consistent with E_1

Conditioning

- If q comes from p by conditioning on E, then $q(E) = 1$
- If $p(E_1) = 1$ then $p(E_1 \mid E_2) = 1$ for any E_2 consistent with E_1
- If $p(E_1) = 1$ then $p(A \mid E_2)$ is undefined whenever E_2 is inconsistent with E_1, since $p(E_2) = 0$

Updating probabilities

Orthodox Bayesian Policy

- accept as admissible input only propositions;
- as response to such an input the only admissible change is conditioning the prior on the proposition in question.

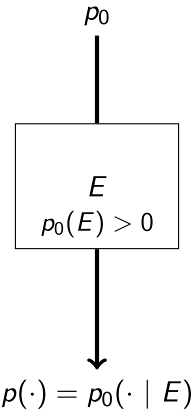
Updating probabilities

Orthodox Bayesian Policy

- accept as admissible input only propositions;
- as response to such an input the only admissible change is conditioning the prior on the proposition in question.

Departing from a (orthodox) Bayesian policy:

1. accept as admissible a wider variety of inputs (e.g. expected values);
2. an admissible response to such an input can be a change in the prior that is not the result of conditioning;
3. an admissible response to such an input may be non-unique, that is, the posterior may not be uniquely determined by the prior + input.

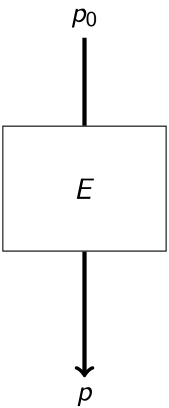

p_0
E
$p_0(E) > 0$
$p(\cdot) = p_0(\cdot \mid E)$

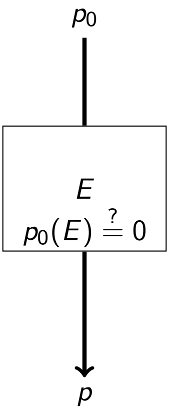
p_0

$$(E_1 : q_1, \ldots, E_k : q_k)$$

$\{E_i\}$ is a partition, $\sum_i q_i = 1$

$$p(\cdot) = \sum_i q_i * p_0(\cdot \mid E_i)$$


p_0
E
p


p_0
E
$p_0(E) \stackrel{?}{=} 0$
p

Conditional Probabilities

What is the probability that a Democrat will be the next president?

Conditional Probabilities

What is the probability that a Democrat will be the next president?

What is the probability that a Democrat will be the next president, *given that a Democrat will be the next president*?

Conditional Probabilities

What is the probability that a Democrat will be the next president?

What is the probability that a Democrat will be the next president, *given that a Democrat will be the next president*?

A. Hájek. *What conditional probability could not be*. Synthese, 137, pp. 273 - 323, 2003.

"When conditional probability is defined by the ratio rule, it has limited expressive capacity. We would like to allow propositions that have been accorded zero probability to serve as conditions for the probability of other propositions. This is impossible when $p(x \mid a)$ is put as $p(a \wedge x)/p(a)$, for it is undefined when $p(a) = 0$."

D. Makinson. *Conditional Probability in the Light of Qualitative Belief Change*. Journal of Philosophical Logic.

Problem: The condition a is consistent but of zero probability (the **critical zone**).

Problem: The condition a is consistent but of zero probability (the **critical zone**).

Solutions:

- Carnap: If $p(x) = 0$, then x is inconsistent.

Problem: The condition a is consistent but of zero probability (the **critical zone**).

Solutions:

- Carnap: If $p(x) = 0$, then x is inconsistent.

  Define $p_a(\cdot)$ as $p(\cdot \mid a)$. By the left projection, $p_a(x) = p(x \mid a)$, then $p_a(\neg a) = p(\neg a \mid a) = 0$ since $p(a)$. Thus, $p_a(\neg a) = 0$ even though $\neg a$ is consistent.

Problem: The condition a is consistent but of zero probability (the **critical zone**).

Solutions:

- Carnap: If $p(x) = 0$, then x is inconsistent.

  Define $p_a(\cdot)$ as $p(\cdot \mid a)$. By the left projection, $p_a(x) = p(x \mid a)$, then $p_a(\neg a) = p(\neg a \mid a) = 0$ since $p(a)$. Thus, $p_a(\neg a) = 0$ even though $\neg a$ is consistent.

- $p(x \mid a) = 1$ for every value x when $p(a) = 0$.

Problem: The condition a is consistent but of zero probability (the **critical zone**).

Solutions:

- Carnap: If $p(x) = 0$, then x is inconsistent.

  Define $p_a(\cdot)$ as $p(\cdot \mid a)$. By the left projection, $p_a(x) = p(x \mid a)$, then $p_a(\neg a) = p(\neg a \mid a) = 0$ since $p(a)$. Thus, $p_a(\neg a) = 0$ even though $\neg a$ is consistent.

- $p(x \mid a) = 1$ for every value x when $p(a) = 0$. *Not very useful.*

Problem: The condition a is consistent but of zero probability (the **critical zone**).

Solutions:

- Carnap: If $p(x) = 0$, then x is inconsistent.

  Define $p_a(\cdot)$ as $p(\cdot \mid a)$. By the left projection, $p_a(x) = p(x \mid a)$, then $p_a(\neg a) = p(\neg a \mid a) = 0$ since $p(a)$. Thus, $p_a(\neg a) = 0$ even though $\neg a$ is consistent.

- $p(x \mid a) = 1$ for every value x when $p(a) = 0$. *Not very useful*.

- $p(x \mid a)$ is the limit of the values of $p(x \mid a')$ for suitable infinite sequence of non-critical approximations a' to a. *Only defined on special domains*.

$$p : \mathcal{L} \times \mathcal{L} \to [0, 1]$$

van Fraassen Axioms:

$vF1$ $\quad p(x, a) = p(x, a')$ whenever $a \equiv a'$

$vF2$ $\quad p_a$ is a one-place Kolmogorov probability function with $p_a(a) = 1$

$vF3$ $\quad p(x \wedge y, a) = p(x, a) * p(y, a \wedge x)$ for all a, x, y

"For 'most' values of the right argument of the two-place function, the left projections should be proper one-place Kolmogorov functions, while in the remaining cases it should be the unit function."

(Positive): when $p(a, \top) > 0$ then p_a is a proper Kolmogorov function.

(Carnap) When a is consistent then $p(a, \top) > 0$.

(Unit) When a is consistent but $p(a, \top) = 0$, then p_a is the unit function.

(HL) When a is consistent but $p(a, \top) = 0$, then p_a is a proper Kolmogorov probability function.

LPS (Lexicographic Probability Space)

A **lexicographic probability space** (LPS) (of length α) is a tuple $(W, \Sigma, \vec{\mu})$ where W is a set of possible worlds, Σ is an algebra over W and $\vec{\mu}$ is a sequence of (finitely/countable additive) probability measures on (W, Σ) indexed by ordinals $< \alpha$.

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$
- E is *assumed*: there exists k such that $\mu_i(E) = 1$ for all $i \leq k$ and $\mu_i(E) = 0$ for all $k < i < n$.

NPS (non-standard probability measures)

$\mathbb{R}^*$ is a *non-Archimedean* field that includes the real numbers as a subfield but also has *infinitesimals*.

For all $b \in \mathbb{R}^*$ such that $-r < b < r$ for some $r \in \mathbb{R}$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. Let $st(b)$ denote the closest standard real to b.

NPS (non-standard probability measures)

$\mathbb{R}^*$ is a *non-Archimedean* field that includes the real numbers as a subfield but also has *infinitesimals*.

For all $b \in \mathbb{R}^*$ such that $-r < b < r$ for some $r \in \mathbb{R}$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. Let $st(b)$ denote the closest standard real to b.

A **nonstandard probability space** (NPS) is a tuple (W, Σ, μ) where W is a set of possible worlds, Σ is an algebra over W and μ assigns to elements of Σ, nonnegative elements of $\mathbb{R}^*$ such that $\mu(W) = 1$, $\mu(E \cup F) = \mu(E) + \mu(F)$ if E and F are disjoint.

J. Halpern. *Lexicographic probability, conditional probability, and nonstandard probability*. Games and Economic Behavior, 68:1, pgs. 155 - 179, 2010.

p_0

E

$p_0(E) > 0$

$$p(\cdot) = p_0(\cdot \mid E)$$

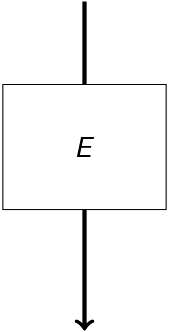
$$p_0$$

$$(E_1 : q_1, \ldots, E_k : q_k)$$

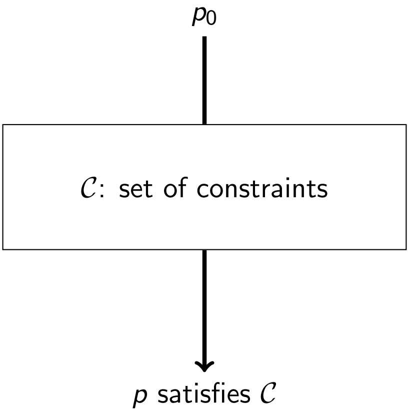
$\{E_i\}$ is a partition, $\sum_i q_i = 1$

$$p(\cdot) = \sum_i q_i * p_0(\cdot \mid E_i)$$

$$p_0(\cdot, \top)$$



$$p(\cdot) = p_0(\cdot, E)$$


p_0
$\mathcal{C}$: set of constraints
p satisfies $\mathcal{C}$

MAXENT

Let us start with the simplest case, where our outcome space, X, contains only a finite number of points, $x_1, x_2, \ldots, x_n$. Then the **entropy** of a probability, p, on this space is:

$$-\sum_i p(x_i) \log p(x_i)$$

and the **information** is the negative of the entropy.

The minimum information or maximum entropy probability is the one which makes the states equiprobable: $p(x_i) = \frac{1}{n}$.

Consider three die x_1, x_2, x_3 and a random variable f such that $f(x_i) = i$.

$$\mathbb{E}[f] = p(x_1)f(x_1) + p(x_2)f(x_2) + p(x_3)f(x_3)$$

What probabilities maximize entropy under the constraint that $\mathbb{E}[f]$ have different values?

MAXENT

$\mathbb{E}[f]$	$p(x_1)$	$p(x_2)$	$p(x_3)$
1	1	0	0
0.1	0.907833	0.084333	0.007834
0.2	0.826297	0.147407	0.026297
⋮	⋮	⋮	⋮
0.8	0.438371	0.323257	0.238271
0.9	0.384586	0.330829	0.284586
2.0	0.333333	0.333333	0.333333
2.1	0.284586	0.330829	0.384586
2.2	0.238372	0.323257	0.438370
⋮	⋮	⋮	⋮
2.8	0.026297	0.147407	0.826296
2.9	0.007834	0.084332	0,907834
3.0	0	0	1

Kullback-Leibler

Suppose that we start with a prior probability, p_0, and move to a posterior p_1 which satisfies certain constraints. The Kullback-Leibler "distance" is:

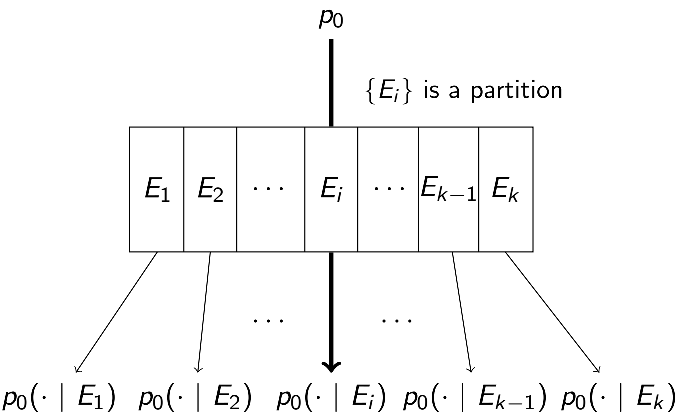
$$I(p_1, p_0) = \sum_i p_1(x_i) \log \frac{p_1(x_i)}{p_0(x_i)}$$

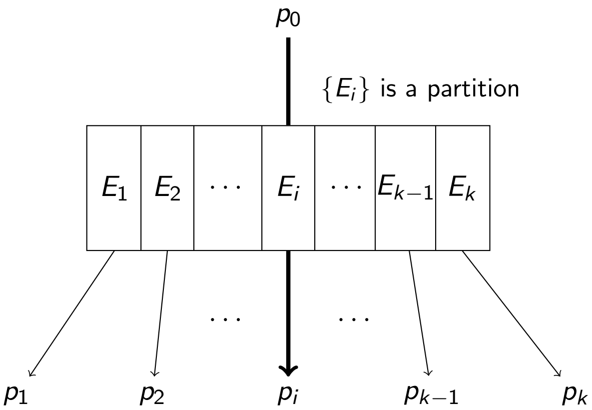
p_0

$\mathcal{C}$: set of constraints

p satisfies $\mathcal{C}$

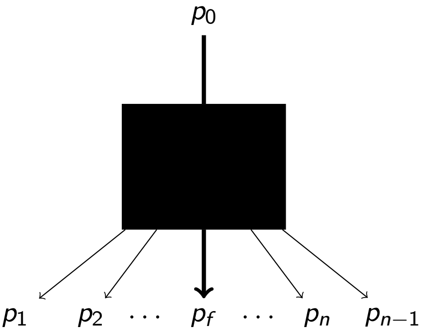
p minimizes $I(p, p_0)$

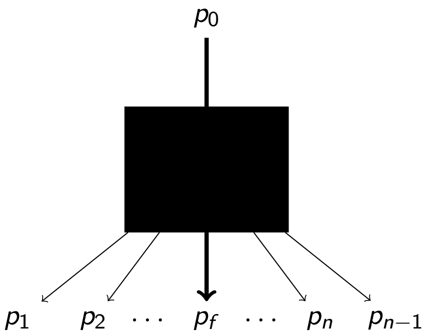

p_0
$\{E_i\}$ is a partition
E_1 E_2 $\cdots$ E_i $\cdots$ E_{k-1} E_k
$\cdots$ $\cdots$
$p_0(\cdot \mid E_1)$ $p_0(\cdot \mid E_2)$ $p_0(\cdot \mid E_i)$ $p_0(\cdot \mid E_{k-1})$ $p_0(\cdot \mid E_k)$


p_0
$\{E_i\}$ is a partition
E_1
E_2
$\cdots$
E_i
$\cdots$
E_{k-1}
E_k
$\cdots$
$\cdots$
p_1
p_2
p_i
p_{k-1}
p_k

Suppose that you are in a learning situation even more amorphous than the kind which motivates Jeffrey's idea. There is no nontrivial partition that you expect with probability one to be sufficient for your belief change....Perhaps you are in a novel situation where you expect the unexpected observational input....You are going to just think about some subject matter and update as a result of your thoughts...I will consider the learning situation a kind of black box and attempt no analysis of its internal structure.

(Skyrms, pg. 96, 97)


p_0
p_1
p_2
$\cdots$
p_f
$\cdots$
p_n
p_{n-1}



(Martingale Property) $\quad p_0(A \mid p_f) = p_f(A)$

It was suggested by Skyrms (1990) that this principle provides a plausible way to distinguish learning situations from situations where one expects probabilities to change for other reasons, such as getting drunk, having a brain lesion or having a dangerously low blood sugar level.

It was suggested by Skyrms (1990) that this principle provides a plausible way to distinguish learning situations from situations where one expects probabilities to change for other reasons, such as getting drunk, having a brain lesion or having a dangerously low blood sugar level.

Huttegger develops an account in which the reflection principle is a necessary condition for a black-box probability update to count as a *genuine learning experience*.

Simon Huttegger. *Learning Experiences and the Value of Knowledge*. Philosophical Studies, 2013.

The Value of Knowledge

Why is it better to make a "more informed" decision?
Suppose that you can either choose know, or perform a costless experiment and make the decision later. What should you do?

I. J. Good. *On the principle of total evidence*. British Journal for the Philosophy of Science, 17, pgs. 319 - 321, 1967.

"Never decide today what you might postpone until tomorrow in order to learn something new"

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i)U(A \;\&\; K_i)$$

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i)U(A \ \& \ K_i)$$

Then,

$$U(\text{Choose now}) = \max_j \sum_i p(K_i)U(A_j \ \& \ K_i)$$

$$= \max_j \sum_k \sum_i p(K_i)p(e_k \mid K_i)U(A_j \ \& \ K_i)$$

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i)U(A \;\&\; K_i)$$

Then,

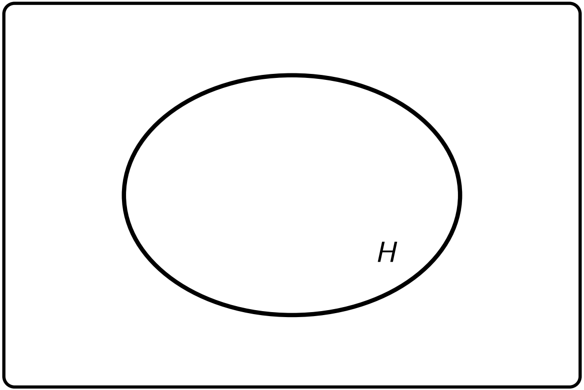
$$U(\text{Choose now}) = \max_j \sum_i p(K_i)U(A_j \;\&\; K_i)$$

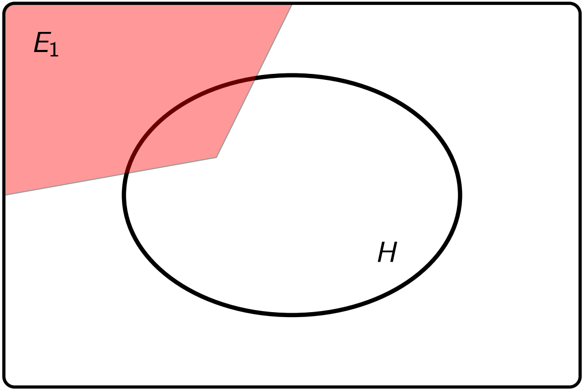
$$= \max_j \sum_k \sum_i p(K_i)p(e_k \mid K_i)U(A_j \;\&\; K_i)$$

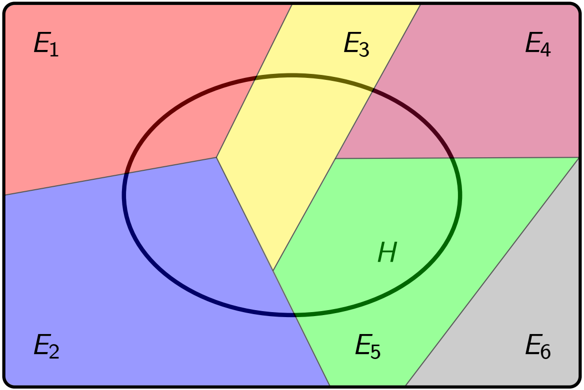
A basic result about probabilities.

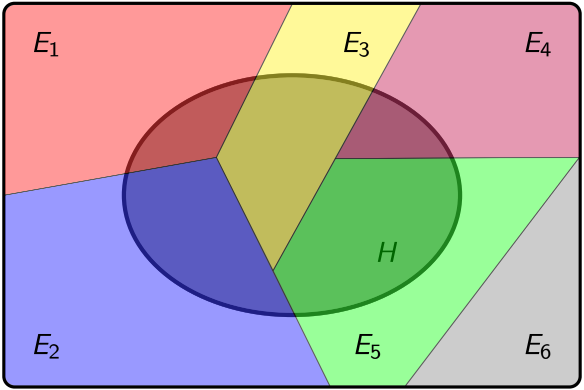
For any finite partition $\{E_i\}$ of the state space and any event H,

$$p(H) = \sum_i p(H \mid E_i)$$

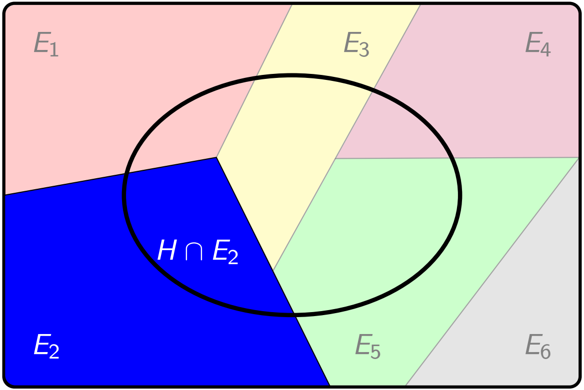

H


E_1
H

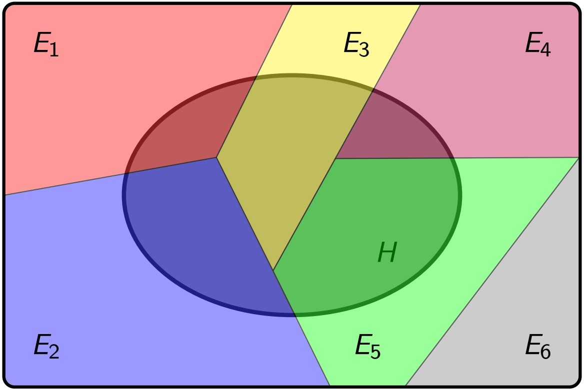

E_1
E_3
E_4
H
E_2
E_5
E_6



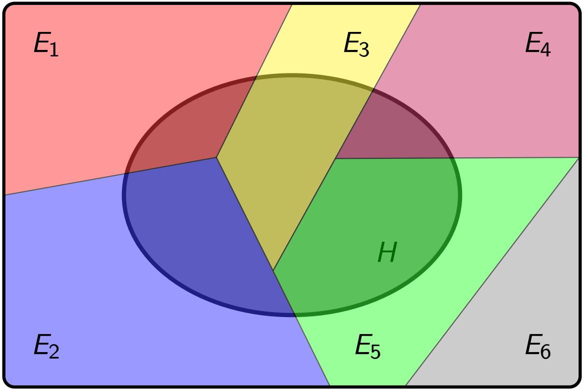
$$p(H) \quad = \quad p(H \cap E_1) + \cdots + p(H \cap E_6)$$



$$p(H) \quad = \quad p(H \cap E_1) + p(H \cap E_2) + \cdots + p(H \cap E_6)$$



$$
\begin{aligned}
p(H) &= p(H \cap E_1) + \cdots + p(H \cap E_6) \\
&= \frac{p(E_1)}{p(E_1)} p(H \cap E_1) + \cdots + \frac{p(E_6)}{p(E_6)} p(H \cap E_6)
\end{aligned}
$$



$$
\begin{aligned}
p(H) &= p(H \cap E_1) + \cdots + p(H \cap E_6) \\
&= \frac{p(E_1)}{p(E_1)} p(H \cap E_1) + \cdots + \frac{p(E_6)}{p(E_6)} p(H \cap E_6) \\
&= \sum_i p(E_i) P(H \mid E_i)
\end{aligned}
$$

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i)U(A \mathbin{\&} K_i)$$

Then,

$$U(\text{Choose now}) = \max_j \sum_i p(K_i)U(A_j \mathbin{\&} K_i)$$

$$= \max_j \sum_i \sum_k p(e_k)p(K_i \mid e_k)U(A_j \mathbin{\&} K_i)$$

Bayes Theorem. $p(K_i|e_j) = p(e_j|K_i)\frac{p(K_i)}{p(e_j)}$

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i)U(A \;\&\; K_i)$$

Then,

$$U(\text{Choose now}) = \max_j \sum_i p(K_i)U(A_j \;\&\; K_i)$$

$$= \max_j \sum_i \sum_k p(e_k)p(e_k \mid K_i)\frac{p(K_i)}{p(e_k)}U(A_j \;\&\; K_i)$$

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i)U(A \,\&\, K_i)$$

Then,

$$U(\text{Choose now}) = \max_j \sum_i p(K_i)U(A_j \,\&\, K_i)$$

$$= \max_j \sum_k \sum_i p(K_i)p(e_k \mid K_i)U(A_j \,\&\, K_i)$$

Choose between n acts $A_1, \ldots, A_n$ (with states K_i) or perform a cost-free experiment with possible results $\{e_k\}$, then decide.

$$EU(A) = \sum_i p(K_i) U(A \ \& \ K_i)$$

Then,

$$U(\text{Choose now}) = \max_j \sum_i p(K_i) U(A_j \ \& \ K_i)$$

$$= \max_j \sum_k \sum_i p(K_i) p(e_k \mid K_i) U(A_j \ \& \ K_i)$$

The value of an informed decision conditional on e:

$$\max_j \sum_i p(K_i \mid e) U(A_j \;\&\; K_i)$$

The value of an informed decision conditional on e:

$$\max_j \sum_i p(K_i \mid e) U(A_j \,\&\, K_i)$$

$$U(\text{Learn, Choose }) \quad = \quad \textstyle\sum_k p(e_k) \max_j \sum_i p(K_i \mid e_k) U(A_j \,\&\, K_i)$$

The value of an informed decision conditional on e:

$$\max_j \sum_i p(K_i \mid e)U(A_j \ \& \ K_i)$$

$$\begin{aligned} U(\text{Learn, Choose }) &= \sum_k p(e_k) \max_j \sum_i p(K_i \mid e_k)U(A_j \ \& \ K_i) \\ &= \sum_k p(e_k) \max_j \sum_i (\frac{p(e_k \mid K_i)p(K_i)}{p(e_k)})U(A_j \ \& \ K_i) \end{aligned}$$

The value of an informed decision conditional on e:

$$\max_j \sum_i p(K_i \mid e)U(A_j \;\&\; K_i)$$

$$\begin{aligned}
U(\text{Learn, Choose }) &= \textstyle\sum_k p(e_k)\max_j \sum_i p(K_i \mid e_k)U(A_j \;\&\; K_i)\\
&= \textstyle\sum_k p(e_k)\max_j \sum_i (\frac{p(e_k \mid K_i)p(K_i)}{p(e_k)})U(A_j \;\&\; K_i)\\
&= \textstyle\sum_k \max_j \sum_i p(e_k \mid K_i)p(K_i)U(A_j \;\&\; K_i)
\end{aligned}$$

The value of an informed decision conditional on e:

$$\max_j \sum_i p(K_i \mid e) U(A_j \,\&\, K_i)$$

$$\begin{aligned} U(\text{Learn, Choose }) &= \textstyle\sum_k p(e_k) \max_j \sum_i p(K_i \mid e_k) U(A_j \,\&\, K_i) \\ &= \textstyle\sum_k p(e_k) \max_j \sum_i \left(\frac{p(e_k \mid K_i) p(K_i)}{p(e_k)}\right) U(A_j \,\&\, K_i) \\ &= \textstyle\sum_k \max_j \sum_i p(e_k \mid K_i) p(K_i) U(A_j \,\&\, K_i) \end{aligned}$$

Compare $\max_j \sum_k \sum_i p(K_i) p(e_k \mid K_i) U(A_j \,\&\, K_i)$ and $\sum_k \max_j \sum_i p(e_k \mid K_i) p(K_i) U(A_j \,\&\, K_i)$

The value of an informed decision conditional on e:

$$\max_j \sum_i p(K_i \mid e) U(A_j \ \& \ K_i)$$

$$\begin{aligned} U(\text{Learn, Choose }) &= \textstyle\sum_k p(e_k) \max_j \sum_i p(K_i \mid e_k) U(A_j \ \& \ K_i) \\ &= \textstyle\sum_k p(e_k) \max_j \sum_i (\frac{p(e_k \mid K_i) p(K_i)}{p(e_k)}) U(A_j \ \& \ K_i) \\ &= \textstyle\sum_k \max_j \sum_i p(e_k \mid K_i) p(K_i) U(A_j \ \& \ K_i) \end{aligned}$$

Compare $\max_j \sum_k \sum_i p(K_i) p(e_k \mid K_i) U(A_j \ \& \ K_i)$ and $\sum_k \max_j \sum_i p(e_k \mid K_i) p(K_i) U(A_j \ \& \ K_i)$

$\sum_k \max_j g(k, j)$ is greater than or equal to $\max_j \sum_k g(k, j)$, so the second is greater than or equal to the first.

(M) implies value of knowledge

Suppose that (M) holds.

(M) implies value of knowledge

Suppose that (M) holds.
Then (assuming that each $p(p_f)$ is positive) your value for choosing an act now is

(M) implies value of knowledge

Suppose that (M) holds.
Then (assuming that each $p(p_f)$ is positive) your value for choosing an act now is

$$\max_j \sum_i p(K_i)u(A_j \& K_i) = \max_j \sum_i \sum_f p(K_i \mid p_f)p(p_f)u(A_j \& K_i)$$

(M) implies value of knowledge

Suppose that (M) holds.
Then (assuming that each $p(p_f)$ is positive) your value for choosing an act now is

$$\max_j \sum_i p(K_i)u(A_j \& K_i) = \max_j \sum_i \sum_f p(K_i \mid p_f)p(p_f)u(A_j \& K_i)$$

$$= \max_j \sum_f \sum_i p_f(K_i)p(p_f)u(A_j \& K_i)$$

(M) implies value of knowledge

Suppose that (M) holds.
Then (assuming that each $p(p_f)$ is positive) your value for choosing an act now is

$$\max_j \sum_i p(K_i)u(A_j \& K_i) = \max_j \sum_i \sum_f p(K_i \mid p_f)p(p_f)u(A_j \& K_i)$$

$$= \max_j \sum_f \sum_i p_f(K_i)p(p_f)u(A_j \& K_i)$$

The value of choosing after the learning experience is:

$$\sum_f p(p_f) \max_j \sum_i p_f(K_i)u(A_j \& K_i)$$

(M) implies value of knowledge

Suppose that (M) holds.
Then (assuming that each $p(p_f)$ is positive) your value for choosing an act now is

$$\max_j \sum_i p(K_i)u(A_j \& K_i) = \max_j \sum_i \sum_f p(K_i \mid p_f)p(p_f)u(A_j \& K_i)$$

$$= \max_j \sum_f \sum_i p_f(K_i)p(p_f)u(A_j \& K_i)$$

The value of choosing after the learning experience is:

$$\sum_f p(p_f) \max_j \sum_i p_f(K_i)u(A_j \& K_i)$$

The latter term cannot be less than the former term on general mathematical grounds.

- The experiment is assumed to be essentially costless;
- You know that you are an expected utility maximizer and that you will be one after learning the true member of the partition.
- In the classical theorem you know that you will update by conditioning; in Skyrms' extension, you know that you will honor the martingale principle.
- By working within Savages decision theory, the states and acts are probabilistically independent (choosing an act does not give any information about the state).

- The states, acts and utilities are the same before and after the learning experience.

- Having the learning experience does not by itself alter your probabilities for states of the world (although the outcomes of the experience usually do); the learning experience and the states of the world are probabilistically independent.

...the martingale principle should not be applied to belief changes in epistemologically defective situations. In situations of memory loss, of being brainwashed or being under the influence of drugs, (M) should obviously not hold. If you believe that in an hour you will think you can fly because you're about to consume some funny looking pills, then you should not already now have that belief.

So, the martingale principle is claimed to apply if you learn something in the black-box, but not if you learn nothing or other things happen besides learning.

A genuine learning situation is partially characterized in the following way:

Postulate. If a belief change from p to $\{pf\}$ constitutes a genuine learning situation, then

$$\sum_f p(p_f) \max_j \sum_i p_f(K_i) u(A_j \& K_i) \geq \max \sum_i p(K_i) u(A_j \& K_i)$$

for all utility values $u(A_j \& K_i)$ with strict inequality unless the same act maximizes expected utility irrespective of which of the p_f occurs.

If a belief change leads you to foreseeably make worse choices than you could already make now in some decision situations, then it cannot be a pure learning experience. Perhaps you are bolder after having taken those funny looking pills, for example. From your current perspective, this might help you in some decision problems, but it will be harmful in others.

Taking Stock

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

Taking Stock

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)
- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

Taking Stock

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)
- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...
- *How* did you find out that φ?
  - Directly observed φ
  - Indirectly observed φ
  - Told 'φ' (by an epistemic peer, by an expert, by a trusted individual)
  - ...

Learn that φ
Suppose that φ

$$K_0 \quad \Longrightarrow \quad K_t = K_0 * \varphi$$

$$\updownarrow \qquad\qquad\qquad \updownarrow$$

Learning experience

$$p_0 \quad \Longrightarrow \quad p_t = ???$$

Learn that φ

Suppose that φ

$$K_0 \quad \Longrightarrow \quad K_t = K_0 * \varphi$$

$$\updownarrow \qquad\qquad\qquad \updownarrow$$

Learning experience

$$p_0 \quad \Longrightarrow \quad p_t = ???$$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

The Nihilistic proposal: "...no explication of belief is possible within the confines of the probability model."

Preface Paradox

D. Makinson. *The Paradox of the Preface*. Analysis, 25, 205 - 207, 1965.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

$B_A(\neg(s_1 \wedge s_2 \wedge \cdots \wedge s_n))$

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

$B_A(\neg(s_1 \wedge s_2 \wedge \cdots \wedge s_n))$

But $\{s_1, \ldots, s_n, \neg(s_1 \wedge \cdots \wedge s_n)\}$ is logically inconsistent.

Preface Paradox

A philosopher who asserts "all of my present philosophical positions are correct" would be regarded as rash and over-confident

A philosopher who asserts "at least some of my present philosophical beliefs will turn out to be incorrect" is simply being sensible and honest.

Preface Paradox

1. each belief from the set $\{s_1, \ldots, s_n, s_{n+1}\}$ is rational
2. the set $\{s_1, \ldots, s_n, s_{n+1}\}$ of beliefs is rational.

1. does not necessarily imply 2.

Preface Paradox: The Problem

"The author of the book is being rational even though inconsistent. More than this: he is being rational even though he believes each of a certain collection of statements, which *he knows* are logically incompatible....this appears to present a living and everyday example of a situation which philosophers have commonly dismissed as absurd; that it is sometimes rational to hold incompatible beliefs."

D. Makinson. *The Paradox of the Preface*. Analysis, 25, 205 - 207, 1965.

H. Leitgeb. *The Review Paradox: On the Diachronic Costs of Not Closing Rational Belief Under Conjunction*. Nous, 2013.

Bel_t is the set of propositions believed at time t

P_t is the agent's degree of belief function at time t

$t' > t$

P1 If the degrees of belief that the agents assigns to two propositions are identical then either the agent believes both of them or neither of them.

For all X, Y: if $P_t(X) = P_t(Y)$, then $Bel_t(X)$ iff $Bel_t(Y)$.

P2 If the agent already believes X, then updating on the piece of evidence X does not change her system of (all-or-nothing) beliefs at all.

For all X: if the evidence that the agent obtains between t and $t' > t$ is the proposition X, but it holds already that $Bel_t(X)$, then for all Y:

$$Bel_{t'}(Y) \text{ iff } Bel_t(Y)$$

P3 When the agent learns, this is captured probabilistically by conditionalization.

For all X (with $P_t(X) > 0$): if the evidence that the agent obtains between t and $t' > t$ is the proposition X, but it holds already that $Bel_t(X)$, then for all Y:

$$P_{t'}(Y) = P_t(Y \mid X)$$

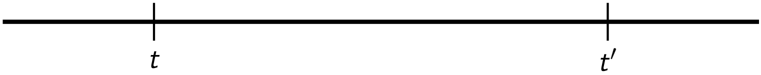
Assume $Bel_t(A), Bel_t(B)$ but not $Bel_t(A \cap B)$

- Suppose that the agent receive A as evidence.
- $P_{t'}(B) = P_t(B \mid A) = P_t(A \cap B \mid A) = P_{t'}(A \cap B)$.
- By P1, the agent must have the same doxastic attitude towards B and $A \cap B$.
- By P2, the agent's attitude towards B and $A \cap B$ must be the same at t' as at t.
- But, $Bel_t(B)$ and not $Bel_t(A \cap B)$

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$



Assumption

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

t — Receives evidence A — t'

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

By P3

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

$Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$

t — Receives evidence A — t'

By P1

$Bel_t(A)$, $Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

$Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$

t — Receives evidence A — t'

$Bel_t(A)$ iff $Bel_{t'}(A)$

$Bel_t(B)$ iff $Bel_{t'}(B)$

$Bel_t(A \cap B)$ iff $Bel_{t'}(A \cap B)$

By P2

$Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

$Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$

t Receives evidence A t'

$Bel_t(A)$ iff $Bel_{t'}(A)$

$Bel_t(B)$ iff $Bel_{t'}(B)$

$Bel_t(A \cap B)$ iff $Bel_{t'}(A \cap B)$

$Bel_t(B)$ iff $Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$ iff $Bel_t(A \cap B)$

Lottery Paradox

H. Kyburg. *Probability and the Logic of Rational Belief*. Wesleyan University Press, 1961.

G. Wheeler. *A Review of the Lottery Paradox*. Probability and Inference: Essays in honor of Henry E. Kyburg, Jr., College Publications, 2007.

Lottery Paradox

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

"Surely if a sheer probability is ever sufficient to warrant the acceptance of a hypothesis, this is a case"

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

"Surely if a sheer probability is ever sufficient to warrant the acceptance of a hypothesis, this is a case"

For each lottery ticket t_i $(i = 1, \ldots, 1000000)$, the agent believes that t_i will loose $B_A(\neg\text{`}t_i \text{ will win'})$

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

So, the conjunction $\bigwedge_{i=1}^{1000000}$ 't_i will not win' should be accepted. That is, the agent should rationally accept that *no lottery ticket will win*.

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

So, the conjunction $\bigwedge_{i=1}^{1000000}$ 't_i will not win' should be accepted. That is, the agent should rationally accept that *no lottery ticket will win*.

But, this is a fair lottery, so at least one ticket is *guaranteed* to win!

The Lottery Paradox

Kyburg: The following are inconsistent,

1. It is rational to accept a proposition that is very likely true,
2. It is not rational to accept a propositional that you are aware is inconsistent
3. It is rational to accept a proposition P and it is rational to accept another proposition P' then it is rational to accept $P \wedge P'$