Dynamics of Inductive Inference in a Unified Model

Itzhak Gilboa, Larry Samuelson, and David Schmeidler

September 13, 2011

Motivation

- September 16, 2001

  What will the DJIA be?

Motivation

- September 16, 2001
  What will the DJIA be?
- September 15, 2008
  “The models do not apply”

Modes of Reasoning

- Bayesian

  Prior on all states; Bayesian updating

Modes of Reasoning

- Bayesian
  Prior on all states; Bayesian updating
- Case-Based
  Analogies; similarities

Modes of Reasoning

- Bayesian
  Prior on all states; Bayesian updating
- Case-Based
  Analogies; similarities
- Rule-Based
  Regularities; deduction, contrapositives...

Prevalence

- Case-based: universal; cats do it

Prevalence

- Case-based: universal; cats do it
- Rule-based: cognitively more demanding

Prevalence

- Case-based: universal; cats do it
- Rule-based: cognitively more demanding
- Bayesian: tends to be difficult; some inference (such as what information I could have gotten but didn't) are quite common

History in Research

- Rule-based: the oldest
  Formal logic, dates back to the Greeks

History in Research

- Rule-based: the oldest
  Formal logic, dates back to the Greeks
- Bayesian: 17th-18th centuries
  Attributed to Bayes, 1763

History in Research

- Rule-based: the oldest
  Formal logic, dates back to the Greeks
- Bayesian: 17th-18th centuries
  Attributed to Bayes, 1763
- Case-based: the latest to be studied academically
  Schank, 1986

Goals

- Develop a model that unifies these modes of reasoning

Goals

- Develop a model that unifies these modes of reasoning
- This would allow

  Comparing them

  Delineating their scope of applicability

  Studying hybrid modes of reasoning

  Studying the dynamics of reasoning

General Model

The primitives are:

- X – a set of *characteristics* that may be observed

General Model

The primitives are:

- X – a set of *characteristics* that may be observed
- Y – a set of *outcomes* that are to be predicted

$$0 < |X|, |Y| < \infty$$

General Model

The primitives are:

- X – a set of *characteristics* that may be observed
- Y – a set of *outcomes* that are to be predicted

$$0 < |X|, |Y| < \infty$$

- $\Omega = (X \times Y)^{\infty}$ – the set of *states of the world*

General Model

The primitives are:

- X – a set of *characteristics* that may be observed
- Y – a set of *outcomes* that are to be predicted

$$0 < |X|, |Y| < \infty$$

- $\Omega = (X \times Y)^{\infty}$ – the set of *states of the world*
- $\mathcal{A} \subset 2^{\Omega}$ – the σ-algebra of *conjectures*

Some more notation

- For a state $\omega \in \Omega$ and a period t, there a *history* up to period t

$$h_t(\omega) = (\omega(0), \ldots, \omega(t-1), \omega_x(t))$$

and its associated *event*

$$[h_t] = \{\omega \in \Omega \mid (\omega(0), \ldots, \omega(t-1), \omega_x(t)) = h_t\}$$

Some more notation

- For a state $\omega \in \Omega$ and a period t, there a *history* up to period t

$$h_t(\omega) = (\omega(0), \ldots, \omega(t-1), \omega_x(t))$$

and its associated *event*

$$[h_t] = \{\omega \in \Omega \mid (\omega(0), \ldots, \omega(t-1), \omega_x(t)) = h_t\}$$

- For a history h_t and a subset of outcomes $Y' \subset Y$ define the event

$$\left[h_t, Y'\right] = \left\{\omega \in [h_t] \,\middle|\, \omega_y(t) \subset Y'\right\}$$

namely, that h_t occurs and results in an outcome in Y'.

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*
- Inference is driven by a σ-additive measure

$$\phi : \mathcal{E} \rightarrow \mathbb{R}_+$$

measuring the degree of belief that the agent has in sets of conjectures.

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*
- Inference is driven by a σ-additive measure

$$\phi : \mathcal{E} \to \mathbb{R}_+$$

  measuring the degree of belief that the agent has in sets of conjectures.

- In the present model, $\phi(\mathcal{C})$ will not change with history.

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*
- Inference is driven by a σ-additive measure

$$\phi : \mathcal{E} \rightarrow \mathbb{R}_+$$

  measuring the degree of belief that the agent has in sets of conjectures.
- In the present model, $\phi(\mathcal{C})$ will not change with history.
- The only inference engine will be pseudo-Bayesian
  Conjectures A that are proven inconsistent with h_t will be discarded

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*
- Inference is driven by a σ-additive measure

$$\phi : \mathcal{E} \to \mathbb{R}_+$$

measuring the degree of belief that the agent has in sets of conjectures.

- In the present model, $\phi(\mathcal{C})$ will not change with history.
- The only inference engine will be pseudo-Bayesian
  Conjectures A that are proven inconsistent with h_t will be discarded
- One may with to make ϕ dependent on history h_t

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*
- Inference is driven by a σ-additive measure

$$\phi : \mathcal{E} \rightarrow \mathbb{R}_+$$

measuring the degree of belief that the agent has in sets of conjectures.

- In the present model, $\phi(\mathcal{C})$ will not change with history.
- The only inference engine will be pseudo-Bayesian
Conjectures A that are proven inconsistent with h_t will be discarded
- One may with to make ϕ dependent on history h_t
- But there's no need to do that.

The Credence Function

- There is a σ-algebra $\mathcal{E}$ of *sets of conjectures*
- Inference is driven by a σ-additive measure

$$\phi : \mathcal{E} \to \mathbb{R}_+$$

  measuring the degree of belief that the agent has in sets of conjectures.

- In the present model, $\phi(\mathcal{C})$ will not change with history.
- The only inference engine will be pseudo-Bayesian
  Conjectures A that are proven inconsistent with h_t will be discarded
- One may with to make ϕ dependent on history h_t
- But there's no need to do that.
- Convention: $\phi(\mathcal{E}) = 1$

Reasoning by Conjectures

- Given history h_t, all conjectures A such that

$$A \cap [h_t] = \varnothing$$

are *refuted* and should be discarded.

Reasoning by Conjectures

- Given history h_t, all conjectures A such that

$$A \cap [h_t] = \varnothing$$

are *refuted* and should be discarded.

- Conjectures A such that

$$A \cap [h_t] = [h_t, Y]$$

say nothing and are *irrelevant*.

How likely is a set of outcomes?

- Given history h_t, how much credence does ϕ lend to each outcome? Or to each set of outcomes?

How likely is a set of outcomes?

- Given history h_t, how much credence does ϕ lend to each outcome? Or to each set of outcomes?
- For $Y' \subsetneq Y$ define

$$\mathcal{A}(h_t, Y') = \left\{ A \in \mathcal{A} \,\middle|\, \varnothing \neq A \cap [h_t] \subset \left[h_t, Y' \right] \right\}$$

which is the class of conjectures that

have not been refuted by h_t

predict that the outcome will be in Y' (hence relevant)

How likely is a set of outcomes?

- Given history h_t, how much credence does ϕ lend to each outcome? Or to each set of outcomes?
- For $Y' \subsetneq Y$ define

$$\mathcal{A}(h_t, Y') = \left\{ A \in \mathcal{A} \,\middle|\, \varnothing \neq A \cap [h_t] \subset \left[h_t, Y'\right] \right\}$$

which is the class of conjectures that

have not been refuted by h_t

predict that the outcome will be in Y' (hence relevant)

- Their weight

$$\phi(\mathcal{A}(h_t, Y'))$$

is the degree of support for the claim that the next observation will be in Y'.

A bit more notation

- Since we have a special interest in subsets of conjectures, define, for $\mathcal{D} \subset \mathcal{A}$,

A bit more notation

- Since we have a special interest in subsets of conjectures, define, for $\mathcal{D} \subset \mathcal{A}$,
- The set of conjectures *in* $\mathcal{D}$ that are unrefuted and predict and outcome in $Y' \subsetneqq Y$

$$\mathcal{D}(h_t, Y') = \left\{ A \in \mathcal{D} \,\middle|\, \varnothing \neq A \cap [h_t] \subset \left[h_t, Y' \right] \right\}$$

A bit more notation

- Since we have a special interest in subsets of conjectures, define, for $\mathcal{D} \subset \mathcal{A}$,
- The set of conjectures *in* $\mathcal{D}$ that are unrefuted and predict and outcome in $Y' \subsetneq Y$

$$\mathcal{D}(h_t, Y') = \left\{ A \in \mathcal{D} \,\middle|\, \varnothing \neq A \cap [h_t] \subset \left[h_t, Y'\right] \right\}$$

- Also, it will be useful to have a notation for the total weight of all conjectures *in* $\mathcal{D}$ that are unrefuted and relevant:

$$\phi(\mathcal{D}(h_t)) = \phi\left(\cup_{Y' \subsetneq Y} \mathcal{D}(h_t, Y')\right)$$

Special Case 1: Bayesian

- The set of Bayesian conjectures:

$$\mathcal{B} = \{\{\omega\} \,|\omega \in \Omega\} \subset \mathcal{A}$$

Special Case 1: Bayesian

- The set of Bayesian conjectures:

$$\mathcal{B} = \{\{\omega\} \, | \omega \in \Omega\} \subset \mathcal{A}$$

- Given a probability p on Ω, one may define

$$\phi_p\left(\{\{\omega\} \, | \, \omega \in A\}\right) = p\left(A\right)$$

and get, for every h_t and every $Y' \subsetneqq Y$,

$$p\left(Y' | \, [h_t]\right) \propto \phi_p(\mathcal{A}(h_t, Y'))$$

Special Case 2: Case-Based

- Consider a simple case-based model of prediction. For a similarity function

$$s : X \times X \rightarrow \mathbb{R}_+$$

define the aggregate similarity for an outcome $y \in Y$

$$S(h_t, y) = \sum_{i=0}^{t-1} \beta^{t-i} s(\omega_x(i), \omega_x(t)) \mathbf{1}_{\{\omega_y(i)=y\}}$$

Special Case 2: Case-Based

- Consider a simple case-based model of prediction. For a similarity function

$$s : X \times X \rightarrow \mathbb{R}_+$$

define the aggregate similarity for an outcome $y \in Y$

$$S(h_t, y) = \sum_{i=0}^{t-1} \beta^{t-i} s(\omega_x(i), \omega_x(t)) \mathbf{1}_{\{\omega_y(i)=y\}}$$

- This is equivalent to kernel classification (with similarity playing the role of the kernel).

Special Case 2: Case-Based

- Consider a simple case-based model of prediction. For a similarity function

$$s : X \times X \rightarrow \mathbb{R}_+$$

define the aggregate similarity for an outcome $y \in Y$

$$S(h_t, y) = \sum_{i=0}^{t-1} \beta^{t-i} s(\omega_x(i), \omega_x(t)) \mathbf{1}_{\{\omega_y(i)=y\}}$$

- This is equivalent to kernel classification (with similarity playing the role of the kernel).
- More involved case-based reasoning is possible, but this is fine for now.

Case-Based cont.

- The case-based conjectures will be of the form

$$A_{i,t,x,z} = \{\omega \in \Omega \,|\, \omega_x(i) = x, \omega_x(t) = z, \omega_y(i) = \omega_y(t)\}$$

for periods $i < t$ and two characteristics $x, z \in X$.

Case-Based cont.

- The case-based conjectures will be of the form

$$A_{i,t,x,z} = \{\omega \in \Omega \,|\, \omega_x(i) = x, \omega_x(t) = z, \omega_y(i) = \omega_y(t)\}$$

  for periods $i < t$ and two characteristics $x, z \in X$.

- $A_{i,t,x,z}$ can be viewed as predicting
  "in period i we'll observe characteristics x, in period t we'll observe characteristics z, and the outcomes will be identical"

Case-Based cont.

- The case-based conjectures will be of the form

$$A_{i,t,x,z} = \{\omega \in \Omega \,|\omega_x(i) = x, \omega_x(t) = z, \omega_y(i) = \omega_y(t)\}$$

  for periods $i < t$ and two characteristics $x, z \in X$.

- $A_{i,t,x,z}$ can be viewed as predicting
  "in period i we'll observe characteristics x, in period t we'll observe characteristics z, and the outcomes will be identical"

- Or:
  "*IF* we observe characteristics x and z in periods i and t, (resp.) *THEN* we'll observe the same outcomes in these periods."

Case-based cont.

- The set of all case-based conjectures is

$$\mathcal{CB} = \{ A_{i,t,x,z} \mid i < t\,,\ x, z \in X \} \subset \mathcal{A}.$$

Case-based cont.

- The set of all case-based conjectures is

$$\mathcal{CB} = \{ A_{i,t,x,z} \mid i < t \, , \, x, z \in X \} \subset \mathcal{A}.$$

- To embed a similarity model, with $s : X \times X \rightarrow \mathbb{R}_+$ in our model, define

$$\phi_{s,\beta}(\{A_{i,t,x,z}\}) = \beta^{(t-i)} s(x,z)$$

to get

$$S(h_t, y) = \phi_{s,\beta}(\mathcal{A}(h_t, \{y\}))$$

Special Case 3: Rule-Based

- Example: an association rule that says "if $x = 1$ then $y = 0$" ("If two countries are democracies then they do not engage in a war")

Special Case 3: Rule-Based

- Example: an association rule that says "if $x = 1$ then $y = 0$" ("If two countries are democracies then they do not engage in a war")
- can be captured by

$$A = \{ \omega \in \Omega \mid \omega(t) \neq (1,1) \quad \forall t \}$$

Rule-based cont.

- A functional rule that says that "$y = f(x)$"
("The price index increases at the same rate as the quantity of money")

$$A = \{\, \omega \in \Omega \mid \omega_y(t) = f(\omega_x(t)) \quad \forall t \,\}.$$

Rule-based cont.

- A functional rule that says that "$y = f(x)$"
  ("The price index increases at the same rate as the quantity of money")
  $$A = \{\, \omega \in \Omega \mid \omega_y(t) = f(\omega_x(t)) \quad \forall t \,\}.$$
- Similarly, one can bound the value of y by $f(x) \pm \delta$ etc.

Rule-based cont.

- A functional rule that says that "$y = f(x)$"
("The price index increases at the same rate as the quantity of money")
$$A = \{\omega \in \Omega \mid \omega_y(t) = f(\omega_x(t)) \quad \forall t\}.$$
- Similarly, one can bound the value of y by $f(x) \pm \delta$ etc.
- We do not offer a general framework for rules. Any refutable "theory" may be modeled as a conjecture, and we do not expect to exhaust the richness of structure of the theories.

The Main Result – Example

- The year is 1960. The reasoner has to predict, for the next 60 years, whether a war will or will not occur. For simplicity, assume that there are no characteristics to observe and consider a finite horizon. Thus,

$$|X| = 1 \qquad |Y| = 2 \qquad T = 60$$

The Main Result – Example

- The year is 1960. The reasoner has to predict, for the next 60 years, whether a war will or will not occur. For simplicity, assume that there are no characteristics to observe and consider a finite horizon. Thus,

$$|X| = 1 \qquad |Y| = 2 \qquad T = 60$$

- There are many states

$$|\Omega| = 2^{T} = 2^{60}$$

The Main Result – Example

- The year is 1960. The reasoner has to predict, for the next 60 years, whether a war will or will not occur. For simplicity, assume that there are no characteristics to observe and consider a finite horizon. Thus,

$$|X| = 1 \qquad |Y| = 2 \qquad T = 60$$

- There are many states

$$|\Omega| = 2^T = 2^{60}$$

- Out of all conjectures ($|\mathcal{A}| = 2^{2^{60}}$) focus on Bayesian and case-based conjectures:

$$\begin{aligned} |\mathcal{B}| &= 2^T = 2^{60} \\ |\mathcal{CB}| &= \binom{T}{2} = \binom{60}{2} \cong 1800 \end{aligned}$$

Example – cont.

- Assume that the reasoner "gives a chance" to CB reasoning

$$\phi(\mathcal{CB}) = \varepsilon; \qquad \phi(\mathcal{B}) = 1 - \varepsilon$$

and splits the weight ϕ within each class of conjectures uniformly.

Example – cont.

- Assume that the reasoner "gives a chance" to CB reasoning

$$\phi(\mathcal{CB}) = \varepsilon; \qquad \phi(\mathcal{B}) = 1 - \varepsilon$$

and splits the weight ϕ within each class of conjectures uniformly.

- Each Bayesian conjecture gets a weight

$$\frac{1-\varepsilon}{2^T} = \frac{1-\varepsilon}{2^{60}}$$

and each case-based conjectures – a weight

$$\frac{\varepsilon}{\binom{T}{2}} \cong \frac{\varepsilon}{1800}$$

Example – cont.

- Assume that the reasoner "gives a chance" to CB reasoning

$$\phi(\mathcal{CB}) = \varepsilon; \qquad \phi(\mathcal{B}) = 1 - \varepsilon$$

and splits the weight ϕ within each class of conjectures uniformly.

- Each Bayesian conjecture gets a weight

$$\frac{1-\varepsilon}{2^T} = \frac{1-\varepsilon}{2^{60}}$$

and each case-based conjectures – a weight

$$\frac{\varepsilon}{\binom{T}{2}} \cong \frac{\varepsilon}{1800}$$

- Now the year is 2010, that is $t = 50$. There are $2^{T-t} = 2^{10}$ unrefuted Bayesian conjectures, and $t = 50$ case-based ones.

Example – cont.

- Thus, the total weight of Bayesian conjectures still in the game is

$$\phi\left(\mathcal{B}(h_t)\right) = 2^{T-t}\frac{1-\varepsilon}{2^T} < \frac{1}{2^t} = \frac{1}{2^{50}}$$

and the case-based ones have total weight

$$\phi\left(\mathcal{CB}(h_t)\right) = t\frac{\varepsilon}{\binom{T}{2}} \cong 50\frac{\varepsilon}{1800}$$

Example – cont.

- Thus, the total weight of Bayesian conjectures still in the game is

$$\phi\left(\mathcal{B}(h_t)\right) = 2^{T-t}\frac{1-\varepsilon}{2^T} < \frac{1}{2^t} = \frac{1}{2^{50}}$$

and the case-based ones have total weight

$$\phi\left(\mathcal{CB}(h_t)\right) = t\frac{\varepsilon}{\binom{T}{2}} \cong 50\frac{\varepsilon}{1800}$$

- Generally,
  $\phi\left(\mathcal{B}(h_t)\right)$ decreases exponentially in t
  $\phi\left(\mathcal{CB}(h_t)\right)$ decreases polynomially (quadratically) in t

Example – cont.

- Thus, the total weight of Bayesian conjectures still in the game is

$$\phi\left(\mathcal{B}(h_t)\right) = 2^{T-t}\frac{1-\varepsilon}{2^T} < \frac{1}{2^t} = \frac{1}{2^{50}}$$

  and the case-based ones have total weight

$$\phi\left(\mathcal{CB}(h_t)\right) = t\frac{\varepsilon}{\binom{T}{2}} \cong 50\frac{\varepsilon}{1800}$$

- Generally,
  $\phi\left(\mathcal{B}(h_t)\right)$ decreases exponentially in t
  $\phi\left(\mathcal{CB}(h_t)\right)$ decreases polynomially (quadratically) in t
- $\Longrightarrow$ For sufficiently large t, reasoning tends to be mostly case-based. (And any other class of conjectures of polynomial size can beat the Bayesian.)

Assumption 1

- We retain the main assumption that the reasoner gives some weight to the case-based conjectures (or to another polynomial class):

Assumption 1

- We retain the main assumption that the reasoner gives some weight to the case-based conjectures (or to another polynomial class):
- **Assumption 1:** $\phi(\mathcal{B}), \phi(\mathcal{CB}) > 0$.

Assumption 2

- We assume some open-mindedness in the way that the weight $\phi_T(\mathcal{B}_T)$ is split. Uniform means that $\forall\ h_t, h'_t \in H_t$,

$$\frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{B}(h'_t)\right)} = 1$$

Assumption 2

- We assume some open-mindedness in the way that the weight $\phi_T(\mathcal{B}_T)$ is split. Uniform means that $\forall\ h_t, h'_t \in H_t$,

$$\frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{B}(h'_t)\right)} = 1$$

- More generally, we can demand

$$\frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{B}(h'_t)\right)} \leq c$$

  or even let c depend on t, provided that c_t does not increase more than polynomially in t:

Assumption 2

- We assume some open-mindedness in the way that the weight $\phi_T(\mathcal{B}_T)$ is split. Uniform means that $\forall\ h_t, h'_t \in H_t$,

$$\frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{B}(h'_t)\right)} = 1$$

- More generally, we can demand

$$\frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{B}(h'_t)\right)} \leq c$$

  or even let c depend on t, provided that c_t does not increase more than polynomially in t:

- **Assumption 2:** $\exists\ P(t),\ \forall\ t\ \forall\ h_t, h'_t \in H_t$,

$$\frac{\phi(\mathcal{B}(h_t))}{\phi(\mathcal{B}(h'_t))} \leq P(t)$$

Assumption 3

- Finally, the weight of the case-based conjectures is assumed to be proportional to the similarity between the characteristics. Specifically,

Assumption 3

- Finally, the weight of the case-based conjectures is assumed to be proportional to the similarity between the characteristics. Specifically,
- **Assumption 3:** There exists a polynomial $Q(t)$ such that, (1) for every i, i', t, t', x, x' and z, z' with $t - i = t' - i'$, and $t' < t$,

$$\frac{\phi(\{A_{i',t',x',z'}\})}{\phi(\{A_{i,t,x,z}\})} \leq Q(t) \tag{1}$$

and (2) for every t, $x, z \in X$ and $i < i' < t$,

$$\frac{\phi(\{A_{i,t,x,z}\})}{\phi(\{A_{i',t,x,z}\})} \leq Q(t). \tag{2}$$

The Main Result

Theorem

Let Assumptions 1-3 hold. Then at each $\omega \in \Omega$,

$$\lim_{t\to\infty} \frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{CB}(h_t)\right)} = 0.$$

- Thus, a pseudo-Bayesian updating rule drives out Bayesian reasoning.

Bayesian Learning

- How come there is no learning? Wasn't the posterior probability of the true state supposed to increase?

Bayesian Learning

- How come there is no learning? Wasn't the posterior probability of the true state supposed to increase?
- Indeed,

$$\frac{p(\{\omega\})}{p([h_t])}$$

grows exponentially with t.

Bayesian Learning

- How come there is no learning? Wasn't the posterior probability of the true state supposed to increase?
- Indeed,

$$\frac{p(\{\omega\})}{p([h_t])}$$

grows exponentially with t.
- But this is so because the denominator is shrinking.

Bayesian Learning

- How come there is no learning? Wasn't the posterior probability of the true state supposed to increase?
- Indeed,

$$\frac{p(\{\omega\})}{p([h_t])}$$

grows exponentially with t.

- But this is so because the denominator is shrinking.
- That is, precisely for the reason that the entire Bayesian mode of thinking fades away.

Bayesian Learning

- How come there is no learning? Wasn't the posterior probability of the true state supposed to increase?
- Indeed,

$$\frac{p\left(\{\omega\}\right)}{p\left([h_t]\right)}$$

grows exponentially with t.
- But this is so because the denominator is shrinking.
- That is, precisely for the reason that the entire Bayesian mode of thinking fades away.
- This doesn't happen if $\varepsilon = 0$: a committed Bayesian will never see how low are the a priori probabilities of the Bayesian conjectures, because she has no alternative to compare them to.

When is Bayesianism Reasonable?

- Our result depends on Assumption 2, which says that the reasoner doesn't know too much about the process (hence cannot favor some states too much).

When is Bayesianism Reasonable?

- Our result depends on Assumption 2, which says that the reasoner doesn't know too much about the process (hence cannot favor some states too much).
- A counterexample: the reasoner knows that the state is ω, and this happens to be true.

When is Bayesianism Reasonable?

- Our result depends on Assumption 2, which says that the reasoner doesn't know too much about the process (hence cannot favor some states too much).
- A counterexample: the reasoner knows that the state is ω, and this happens to be true.
- Clearly, Assumption 2 is violated.

When is Bayesianism Reasonable?

- Our result depends on Assumption 2, which says that the reasoner doesn't know too much about the process (hence cannot favor some states too much).
- A counterexample: the reasoner knows that the state is ω, and this happens to be true.
- Clearly, Assumption 2 is violated.
- Such a reasoner would have no reason to abandon the Bayesian belief.

Reasonable Bayesianism – cont.

- More generally: the reasoner may know the process up to k parameters

  and k does not grow with t

Reasonable Bayesianism – cont.

- More generally: the reasoner may know the process up to k parameters
  and k does not grow with t
- Example: observing a comet
  knowing that the phenomenon is cyclical.

Reasonable Bayesianism – cont.

- More generally: the reasoner may know the process up to k parameters
  and k does not grow with t
- Example: observing a comet
  knowing that the phenomenon is cyclical.
- Bayesianism will survive if
  The reasoner believes that she knows the process
  She happens to be right.

The IID Case

- A probability measure μ on Σ is a *non-trivial conditionally iid measure* if, for every $x \in X$ there exists $\lambda_x \in \Delta(Y)$ such that (i) for every h_t, the conditional distribution of Y given h_t according to μ is λ_{x_t}; and (ii) λ_x is non-degenerate for every $x \in X$.

The IID Case

- A probability measure μ on Σ is a *non-trivial conditionally iid measure* if, for every $x \in X$ there exists $\lambda_x \in \Delta(Y)$ such that (i) for every h_t, the conditional distribution of Y given h_t according to μ is λ_{x_t}; and (ii) λ_x is non-degenerate for every $x \in X$.
- **Assumption 2':** There exists a non-trivial conditionally iid measure μ such that, for every $A \in \Sigma$

$$\varphi\left(\{\{\omega\} \mid \omega \in A\}\right) = \mu(A)\varphi(\mathcal{B})$$

The IID Case – Result

Theorem

Let Assumptions 1-3 hold. Then

$$\mu\left(\lim_{t\to\infty}\frac{\phi\left(\mathcal{B}(h_t)\right)}{\phi\left(\mathcal{CB}(h_t)\right)}=0\right)=1.$$

How do Case-Based Conjectures Survive?

- Imagine that each conjecture is a consultant.

How do Case-Based Conjectures Survive?

- Imagine that each conjecture is a consultant.
- They sit in a room at $t = 0$ and state predictions A.

How do Case-Based Conjectures Survive?

- Imagine that each conjecture is a consultant.
- They sit in a room at $t = 0$ and state predictions A.
- As history unfolds, the refuted ones are asked to leave.

How do Case-Based Conjectures Survive?

- Imagine that each conjecture is a consultant.
- They sit in a room at $t = 0$ and state predictions A.
- As history unfolds, the refuted ones are asked to leave.
- Case-based consultants are allowed to say "I don't know".
  $A_{2003,2010,x,z}$ says something about $t = 2010$, but nothing about other t's

How do Case-Based Conjectures Survive?

- Imagine that each conjecture is a consultant.
- They sit in a room at $t = 0$ and state predictions A.
- As history unfolds, the refuted ones are asked to leave.
- Case-based consultants are allowed to say "I don't know".
  $A_{2003,2010,x,z}$ says something about $t = 2010$, but nothing about other t's
- Commitment to Bayesianism means that the weight $\phi\left(A_{2003,2010,x,z}\right)$ has to be split among the 2^{58} states in $A_{2003,2010,x,z}$. Most of these will be wrong.

How do Case-Based Conjectures Survive?

- Imagine that each conjecture is a consultant.
- They sit in a room at $t = 0$ and state predictions A.
- As history unfolds, the refuted ones are asked to leave.
- Case-based consultants are allowed to say "I don't know".
  $A_{2003,2010,x,z}$ says something about $t = 2010$, but nothing about other t's
- Commitment to Bayesianism means that the weight $\phi\left(A_{2003,2010,x,z}\right)$ has to be split among the 2^{58} states in $A_{2003,2010,x,z}$. Most of these will be wrong.
- Leaving the case-based consultant in the room is like crediting him with knowing when to remain silent. As if the meta-knowledge (when do I really know something) is another criterion in the selection of consultants.

Comments

- Convergence to an additive probability
  but a frequentist (non-Bayesian) one.

Comments

- Convergence to an additive probability
  but a frequentist (non-Bayesian) one.
- Similar results could apply to families of rule based conjectures
  and may generate non-additive probability.

Comments

- Convergence to an additive probability
  but a frequentist (non-Bayesian) one.

- Similar results could apply to families of rule based conjectures
  and may generate non-additive probability.

- A different interpretation: the result describes the formation of prior probability.
  If one knows how to split weight among states (Laplace?).

Case-Based vs. Rule-Based Dynamics

- The weight of the case-based conjectures is fixed

Case-Based vs. Rule-Based Dynamics

- The weight of the case-based conjectures is fixed
- Each rule (or theory) has a high weight a priori
  - If successful, the reasoner is mostly rule-based
  - If not, the cases are always there

Algorithms

- Often the carrier of credence is not a particular conjecture, but an algorithm to generate one.

Algorithms

- Often the carrier of credence is not a particular conjecture, but an algorithm to generate one.
- Example: OLS

  The particular regression line is not the issue

  It's the method of generating it

Algorithms

- Often the carrier of credence is not a particular conjecture, but an algorithm to generate one.
- Example: OLS

  The particular regression line is not the issue

  It's the method of generating it
- Another version: carriers are classes of conjectures, with maximum likelihood within each one.

Other Directions

- Probabilistic version
  - Rules replaced by distributions
  - Refutation – by likelihood
  - Several ways to proceed

Other Directions

- Probabilistic version
  - Rules replaced by distributions
  - Refutation – by likelihood
  - Several ways to proceed
- Decision theory
  - For example, payoff is only at terminal states
  - One can use Choquet expected utility
  - There could be multiple ϕ's (with maxmin over them?)