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
General Model

The primitives are:

- \( X \) – a set of *characteristics* that may be observed
- \( Y \) – a set of *outcomes* that are to be predicted
- \( T \) – *number of periods*

\[ 0 < |X|, |Y|, T < \infty \]
General Model
The primitives are:

- \( X \) – a set of characteristics that may be observed
- \( Y \) – a set of outcomes that are to be predicted
- \( T \) – number of periods

\[
0 < |X|, |Y|, T < \infty
\]

- \( \Omega = (X \times Y)^T \) – 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
- $T$ – *number of periods*

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

- $\Omega = (X \times Y)^T$ – the set of *states of the world*
- $A = 2^\Omega$ – the set of *hypotheses*
Some more notation

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

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

and its associated event

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

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

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

and its associated event

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

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

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

namely, that $h_t$ occurs and results in an outcome in $Y'$. 
The Credence Function

- Inference is driven by a function

\[ \phi : \mathcal{A} \rightarrow \mathbb{R}_+ \]

measuring the degree of belief that the reasoner has in each hypothesis.
The Credence Function

- Inference is driven by a function

\[ \phi : \mathcal{A} \rightarrow \mathbb{R}_+ \]

measuring the degree of belief that the reasoner has in each hypothesis.

- In the present model, \( \phi(A) \) will not change with history.
The Credence Function

- Inference is driven by a function

\[ \phi : \mathcal{A} \rightarrow \mathbb{R}_+ \]

measuring the degree of belief that the reasoner has in each hypothesis.

- In the present model, \( \phi(A) \) will not change with history.

- The only inference engine will be pseudo-Bayesian

Hypotheses \( A \) that are proven inconsistent with \( h_t \) will be discarded.
The Credence Function

- Inference is driven by a function

\[ \phi : \mathcal{A} \rightarrow \mathbb{R}_+ \]

measuring the degree of belief that the reasoner has in each hypothesis.

- In the present model, \( \phi(A) \) will not change with history.

- The only inference engine will be pseudo-Bayesian

Hypotheses \( A \) that are proven inconsistent with \( h_t \) will be discarded.

- In a realistic model, since

\[ |\mathcal{A}| = 2^{|\Omega|} = 2^{|X| \times |Y|} \]

is large, most hypotheses are not a priori conceived of by the reasoner. This is not a model of bounded rationality.
Conventions

- Since we are interested in classes of hypotheses, it will be useful to define, for $D \subset A$

\[ \phi(D) = \sum_{A \in D} \phi(A) \]
Conventions

- Since we are interested in classes of hypotheses, it will be useful to define, for $\mathcal{D} \subset \mathcal{A}$
  $$\phi(\mathcal{D}) = \sum_{A \in \mathcal{D}} \phi(A)$$

- We will also normalize $\phi$:
  $$\phi(\mathcal{A}) = 1$$
Reasoning by Hypotheses

- Given history $h_t$, all hypotheses $A$ such that

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

are refuted and should be discarded.
Reasoning by Hypotheses

- Given history $h_t$, all hypotheses $A$ such that
  \[ A \cap [h_t] = \emptyset \]
  are \textit{refuted} and should be discarded.

- Hypotheses $A$ such that
  \[ A \cap [h_t] = [h_t, Y] \]
  say nothing and are \textit{irrelevant}. 

How likely is a set of outcomes?

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

$$A(h_t, Y') = \{ A \in A \mid \emptyset \neq A \cap [h_t] \subset [h_t, Y'] \}$$

which is the class of hypotheses 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 $\phi$ lend to each outcome?
  Or to each set of outcomes?
- For $Y' \subsetneq Y$ define

$$ \mathcal{A}(h_t, Y') = \{ A \in \mathcal{A} \mid \emptyset \neq A \cap [h_t] \subset [h_t, Y'] \} $$

which is the class of hypotheses 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 $Y'$. 

Gilboa, Samuelson, and Schmeidler (2010) Dynamics of Inductive Inference in a Unified Model
A bit more notation

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

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

- The set of hypotheses in $\mathcal{D}$ that are unrefuted and predict and outcome in $Y' \subsetneq Y$

\[
\mathcal{D}(h_t, Y') = \{ A \in \mathcal{D} \mid \emptyset \neq A \cap [h_t] \subset [h_t, Y'] \}
\]
A bit more notation

- Since we have a special interest in subsets of hypotheses, define, for $\mathcal{D} \subseteq \mathcal{A}$,
- The set of hypotheses in $\mathcal{D}$ that are unrefuted and predict and outcome in $Y' \subseteq Y$
  \[ \mathcal{D}(h_t, Y') = \left\{ A \in \mathcal{D} \mid \emptyset \neq A \cap [h_t] \subseteq [h_t, Y'] \right\} \]
- Also, it will be useful to have a notation for the total weight of all hypotheses in $\mathcal{D}$ that are unrefuted and relevant:
  \[ \phi(\mathcal{D}(h_t)) = \phi \left( \bigcup_{Y' \subseteq Y} \mathcal{D}(h_t, Y') \right) \]
Special Case 1: Bayesian

- The set of Bayesian hypotheses:

\[ \mathcal{B} = \{ \{\omega\} \mid \omega \in \Omega \} \subset \mathcal{A} \]
Special Case 1: Bayesian

- The set of Bayesian hypotheses:
  \[ B = \{ \{ \omega \} | \omega \in \Omega \} \subset A \]

- Given a probability \( p \) on \( \Omega \), one may define
  \[ \phi_p (\{ \omega \}) = p (\{ \omega \}) \]
  and get, for every \( h_t \) and every \( Y' \subsetneq Y \),
  \[ p (Y' | [h_t]) \propto \phi_p (A(h_t, Y')) \]
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 \( y \in Y \)

\[
S(h_t, y) = \sum_{i=0}^{t-1} \beta^{t-i} s(\omega_x(i), \omega_x(t))1_{\omega_y(i)=y}
\]
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 hypotheses will be of the form

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

for periods \( i < t \leq T \) and two characteristics \( x, z \in X \).
Case-Based cont.

- The case-based hypotheses will be of the form

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

for periods \( i < t \leq T \) and two characteristics \( x, z \in X \).

- \( A_{it,x,z} \) 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 hypotheses will be of the form

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

for periods \( i < t \leq T \) and two characteristics \( x, z \in X \).

- \( A_{it,x,z} \) 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 hypotheses is

\[ \mathcal{CB} = \{ A_{it,x,z} | i < t \leq T, x, z \in X \} \subset \mathcal{A}. \]
Case-based cont.

- The set of all case-based hypotheses is

\[ CB = \{ A_{it,x,z} \mid i < t \leq T, x, z \in X \} \subset A. \]

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

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

and get

\[ S(h_t, y) = \phi_{s,\beta}(A(h_t, \{y\})) \]
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”)

\[ A = f_{\omega_2} \Omega_j(\omega) t_\xi = (1, 1) \]

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 hypothesis, and we do not expect to exhaust the richness of structure of the theories.
The Dynamics of Reasoning

- How does the overall weight assigned to hypotheses change as evidence is accumulated?
The Dynamics of Reasoning

- How does the overall weight assigned to hypotheses change as evidence is accumulated?
- We are interested in asymptotic results, as \( T \to \infty \).
The Dynamics of Reasoning

- How does the overall weight assigned to hypotheses change as evidence is accumulated?
- We are interested in asymptotic results, as $T \to \infty$.
- Since $T < \infty$, we consider a sequence of models.
The Dynamics of Reasoning

- How does the overall weight assigned to hypotheses change as evidence is accumulated?
- We are interested in asymptotic results, as $T \to \infty$.
- Since $T < \infty$, we consider a sequence of models.
- $X$ and $Y$ are fixed throughout.
The Dynamics of Reasoning

- How does the overall weight assigned to hypotheses change as evidence is accumulated?
- We are interested in asymptotic results, as $T \to \infty$.
- Since $T < \infty$, we consider a sequence of models.
- $X$ and $Y$ are fixed throughout.
- For every $T$ we have a new model, with a state space $\Omega_T$. 
The Dynamics of Reasoning

- How does the overall weight assigned to hypotheses change as evidence is accumulated?
- We are interested in asymptotic results, as $T \to \infty$.
- Since $T < \infty$, we consider a sequence of models.
- $X$ and $Y$ are fixed throughout.
- For every $T$ we have a new model, with a state space $\Omega_T$.
- We will define the set of Bayesian and Case-based hypotheses for each $T$

$$B_T, CB_T$$
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. Thus,

\[ |X| = 1 \quad |Y| = 2 \quad T = 60 \]
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. Thus,

\[ |X| = 1 \quad |Y| = 2 \quad T = 60 \]

- 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. Thus,

\[ |X| = 1 \quad |Y| = 2 \quad T = 60 \]

- There are many states \( |\Omega| = 2^T = 2^{60} \)

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

\[ |B| = 2^T = 2^{60} \]

\[ |CB| = \binom{T}{2} = \binom{60}{2} \approx 1800 \]
Example – cont.

- Assume that the reasoner “gives a chance” to CB reasoning

\[
\phi_T(CB_T) = \varepsilon; \quad \phi_T(B_T) = 1 - \varepsilon
\]

and splits the weight \( \phi \) within each class of hypotheses uniformly.
Example – cont.

- Assume that the reasoner “gives a chance” to CB reasoning
  \[ \phi_T(\text{CB}_T) = \varepsilon; \quad \phi_T(\text{B}_T) = 1 - \varepsilon \]

  and splits the weight \( \phi \) within each class of hypotheses uniformly.

- Each Bayesian hypothesis gets a weight
  \[ \frac{1 - \varepsilon}{2^T} = \frac{1 - \varepsilon}{2^{60}} \]

  and each case-based hypotheses – a weight
  \[ \frac{\varepsilon}{\frac{T}{2}} \approx \frac{\varepsilon}{1800} \]
Example – cont.

- Assume that the reasoner “gives a chance” to CB reasoning

\[ \phi_T(CB_T) = \varepsilon; \quad \phi_T(B_T) = 1 - \varepsilon \]

and splits the weight \( \phi \) within each class of hypotheses uniformly.

- Each Bayesian hypothesis gets a weight

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

and each case-based hypotheses – a weight

\[ \frac{\varepsilon}{(T/2)} \approx \frac{\varepsilon}{1800} \]

- Now the year is 2010, that is \( t = 50 \). There are \( 2^{T-t} = 2^{10} \) unfuted Bayesian hypotheses, and \( t = 50 \) case-based ones.
Example – cont.

Thus, the total weight of Bayesian hypotheses still in the game is

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

and the case-based ones have total weight

$$\phi_T(\mathcal{CB}_T(h_t)) = t \frac{\varepsilon}{T \choose 2} \approx 50 \frac{\varepsilon}{1800}$$
Thus, the total weight of Bayesian hypotheses still in the game is

\[ \phi_T (\mathcal{B}_T(h_t)) = 2^{-t} \frac{1 - \varepsilon}{2^T} < \frac{1}{2^t} = \frac{1}{2^{50}} \]

and the case-based ones have total weight

\[ \phi_T (\mathcal{CB}_T(h_t)) = t \frac{\varepsilon}{T} \approx 50 \frac{\varepsilon}{1800} \]

Generally,

\[ \phi_T (\mathcal{B}_T(h_t)) \] decreases exponentially in \( t \)

\[ \phi_T (\mathcal{CB}_T(h_t)) \] decreases polynomially (quadratically) in \( T \)
Example – cont.

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

$$\phi_T(B_T(h_t)) = 2^{T-t} \frac{1 - \varepsilon}{2^T} < \frac{1}{2^t} = \frac{1}{2^{50}}$$

and the case-based ones have total weight

$$\phi_T(CB_T(h_t)) = t \frac{\varepsilon}{(T^2)} \approx 50 \frac{\varepsilon}{1800}$$

- Generally,

$$\phi_T(B_T(h_t))$$ decreases exponentially in $t$
$$\phi_T(CB_T(h_t))$$ decreases polynomially (quadratically) in $T$

- $\implies$ For sufficiently large $t, T$, reasoning tends to be mostly case-based.
  (And any other class of hypotheses of polynomial size can beat the Bayesian.)
Assumption 1

- We retain the main assumption that the reasoner gives some weight to the case-based hypotheses (or to another polynomial class):
Assumption 1

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

- **Assumption 1**: $\phi_T(B_T), \phi_T(CB_T) > \varepsilon$. 
Assumption 1

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

  **Assumption 1:** $\phi_T(B_T), \phi_T(CB_T) > \epsilon$.

- Importantly, $\epsilon$ is independent of $T$. 
Assumption 2

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

$$\frac{\phi_T(A)}{\phi_T(B)} = 1$$
Assumption 2

- We assume some open-mindedness in the way that the weight \( \phi_T(B_T) \) is split. Uniform means that \( \forall A, B \in \mathcal{B}_T, \)

\[
\frac{\phi_T(A)}{\phi_T(B)} = 1
\]

- More generally, we can demand

\[
\frac{\phi_T(A)}{\phi_T(B)} \leq c
\]

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 $\forall A, B \in \mathcal{B}_T$,

$$\frac{\phi_T(A)}{\phi_T(B)} = 1$$

- More generally, we can demand

$$\frac{\phi_T(A)}{\phi_T(B)} \leq c$$

or even let $c$ depend on $T$, provided that $c_T$ does not increase more than polynomially in $T$:

- **Assumption 2:** $\exists P(T), \forall T \forall A, B \in \mathcal{B}_T$,

$$\phi_T(A) \leq P(T)\phi_T(B)$$
Assumption 3

Finally, the weight of the case-based hypotheses is assumed to be proportional to the similarity between the characteristics. Specifically,
Assumption 3

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

**Assumption 3:** \( \exists s : X \times X \rightarrow \mathbb{R}_{++} \exists \beta \in (0, 1] \) such that \((\forall T \exists c_T > 0) \forall i < t < T, x, z \in X, \phi_T(A_{it,x,z}) = c_T \beta^{t-i} s(x, z)\)
Assumption 3

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

- **Assumption 3:** \( \exists \ s : X \times X \to \mathbb{R}_{++} \ \exists \ \beta \in (0, 1] \) such that \( (\forall T \ \exists c_T > 0) \ \forall i < t < T, x, z \in X, \)

\[
\phi_T(A_{it,x,z}) = c_T \beta^{t-i} s(x, z)
\]

- Observe that \( s \) is assumed to be strictly positive. \( c_T \) is a normalization factor.
The Main Result

Proposition

Let Assumptions 1-3 hold. Then for every $\alpha, \delta > 0$ there exists $T_0$ such that, for every $T > \frac{1}{\alpha} T_0$, every $t \geq \alpha T$, and every history $h_t$,

$$\frac{\phi_T (B_T(h_t))}{\phi_T (A_T \setminus B_T(h_t))} < \delta.$$

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?

Indeed, $p(f_{\omega|g})$ 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 hypotheses, because she has no alternative to compare them to.
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,

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

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.
- This doesn’t happen if \( \varepsilon = 0 \): a committed Bayesian will never see how low are the a priori probabilities of the Bayesian hypotheses, 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 $\omega$, 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 $\omega$, 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 $\omega$, 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.
How do Case-Based Hypotheses Survive?

- Imagine that each hypothesis is a consultant.
How do Case-Based Hypotheses Survive?

- Imagine that each hypothesis is a consultant.
- They sit in a room at $t = 0$ and state predictions $A$. 
How do Case-Based Hypotheses Survive?

- Imagine that each hypothesis 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 Hypotheses Survive?

- Imagine that each hypothesis 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 Hypotheses Survive?

- Imagine that each hypothesis 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 $\phi(A_{2003, 2010, x, z})$ 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 Hypotheses Survive?

- Imagine that each hypothesis 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 hypotheses 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 hypotheses 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

- In the paper we play with some calculations
Case-Based vs. Rule-Based Dynamics

- In the paper we play with some calculations
- The weight of the case-based hypotheses is fixed
Case-Based vs. Rule-Based Dynamics

- In the paper we play with some calculations
- The weight of the case-based hypotheses 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 hypothesis, but an algorithm to generate one.
Algorithms

- Often the carrier of credence is not a particular hypothesis, 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 hypothesis, 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 hypotheses, 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 $\phi$’s (with maxmin over them?)