Douven’s Model vs Our Bayesian Model of “Abductive Updating”

Branden Fitelson
August 2, 2022

\[\text{In}[-\beta] = \text{PrSAT}`

Here is a Bayesian model of an agent’s priors \(c_{r_t}(\ast)\), which are such that:

1. \(c_{r_t}(\ast)\) is regular,
2. \(E\) confirms \(H\): \(c_{r_t}(H \mid E) > c_{r_t}(H \mid \neg E)\),
3. \(ERA\) holds: \(c_{r_t}(E \mid H \& A) > c_{r_t}(E \mid H \& \neg A)\),
   - Here, \(A\) is interpreted as asserting that \(H\) is the best (or only) explanation of \(E\).
4. \(E\) is \(a\ priori\) irrelevant to \(A\): \(c_{r_t}(A \mid E) = c_{r_t}(A \mid \neg E)\),
5. \(A\) is \(a\ priori\) irrelevant to \(H\): \(c_{r_t}(H \mid A) = c_{r_t}(H \mid \neg A)\), and
6. \(c_{r_t}(H) = c_{r_t}(E) = c_{r_t}(A) = \frac{1}{2}.

\[\text{In}[-\beta] = \text{prior} = \text{PrSAT}[\{
\begin{align*}
\text{Pr}[H \mid E] & > \text{Pr}[H \mid \neg E], \\
\text{Pr}[H \mid E \& A] & > \text{Pr}[H \mid E \& \neg A], \\
\text{Pr}[A \mid E] & = \text{Pr}[A \mid \neg E], \\
\text{Pr}[H \mid A] & = \text{Pr}[H \mid \neg A], \\
\text{Pr}[H] & = \frac{1}{2}, \text{Pr}[E] = \frac{1}{2}, \text{Pr}[A] = \frac{1}{2}
\end{align*}
\}, \text{Probabilities} \rightarrow \text{Regular}, \text{BypassSearch} \rightarrow \text{True}]

\[\text{Out}[-\beta] = \{\text{A} \rightarrow \{a_2, a_5, a_6, a_8\}, \text{E} \rightarrow \{a_3, a_5, a_7, a_8\}, \text{H} \rightarrow \{a_4, a_6, a_7, a_8\}, \Omega \rightarrow \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}\}

\{a_1 \rightarrow \frac{3}{32}, a_2 \rightarrow \frac{7}{32}, a_3 \rightarrow \frac{5}{32}, a_4 \rightarrow \frac{5}{32}, a_5 \rightarrow \frac{1}{32}, a_6 \rightarrow \frac{1}{32}, a_7 \rightarrow \frac{3}{32}, a_8 \rightarrow \frac{7}{32}\}
If we look at this agent's credences over the sub-algebra \{H,E\}, we have the following (coarse-graining of their) prior.

\[
\begin{array}{c|c|c|c}
A & E & H & \text{cr}_{t_0} \\
\hline
T & T & T & \frac{7}{32} \\
T & T & F & \frac{1}{32} \\
T & F & T & \frac{1}{32} \\
T & F & F & \frac{7}{32} \\
F & T & T & \frac{3}{32} \\
F & T & F & \frac{5}{32} \\
F & F & T & \frac{5}{32} \\
F & F & F & \frac{3}{32} \\
\end{array}
\]

Here's what the Bayesian evolution of this prior looks like, first at \(t_0\) (before anything is learned), then at \(t_1\) (when \(E\) is learned), and finally at \(t_2\) (when \(A\) is learned).

\[
\text{EvaluateProbability}[[\text{Pr}[A \land E \land H], \text{Pr}[A \land E \land \neg H], \text{Pr}[\neg A \land E \land H], \text{Pr}[\neg A \land E \land \neg H]], \text{prior}] \\
\]

\[
\begin{array}{c|c|c|c}
A & E & H & \text{cr}_{t_0} \\
\hline
T & T & T & \frac{7}{32} \\
T & T & F & \frac{1}{32} \\
T & F & T & \frac{1}{32} \\
T & F & F & \frac{7}{32} \\
F & T & T & \frac{3}{32} \\
F & T & F & \frac{5}{32} \\
F & F & T & \frac{5}{32} \\
F & F & F & \frac{3}{32} \\
\end{array}
\]

\[
\text{EvaluateProbability}[[\text{Pr}[A \land E \land H | E], \text{Pr}[A \land E \land \neg H | E], \text{Pr}[\neg A \land E \land H | E], \text{Pr}[\neg A \land E \land \neg H | E]], \text{prior}] \\
\]

\[
\begin{array}{c|c|c|c}
A & E & H & \text{cr}_{t_1} \\
\hline
T & T & T & \frac{5}{16} \\
T & T & F & \frac{3}{16} \\
T & F & T & \frac{3}{16} \\
T & F & F & \frac{5}{16} \\
F & T & T & \frac{5}{16} \\
F & T & F & \frac{3}{16} \\
F & F & T & \frac{3}{16} \\
F & F & F & \frac{5}{16} \\
\end{array}
\]

\[
\text{EvaluateProbability}[[\text{Pr}[A \land E \land H | E], \text{Pr}[A \land E \land \neg H | E], \text{Pr}[\neg A \land E \land H | E], \text{Pr}[\neg A \land E \land \neg H | E]], \text{prior}] \\
\]

\[
\begin{array}{c|c|c|c}
A & E & H & \text{cr}_{t_2} \\
\hline
T & T & T & \frac{5}{16} \\
T & T & F & \frac{3}{16} \\
T & F & T & \frac{3}{16} \\
T & F & F & \frac{5}{16} \\
F & T & T & \frac{5}{16} \\
F & T & F & \frac{3}{16} \\
F & F & T & \frac{3}{16} \\
F & F & F & \frac{5}{16} \\
\end{array}
\]
What does Douven’s model of an analogous agent look like? Well, presumably, it starts off the same, but only on the sub-algebra \{\text{H,E}\}, since the agent is not assumed to have \text{A} in their algebra of entertained propositions.

Here, if the agent were to follow Bayes’s Theorem, then they would assign a value $\frac{5}{8}$ to $\text{cr}_{\text{t}_2}(\text{H} \mid \text{E})$.
\[
cr_{t_2}(H \mid E) = \frac{cr_{t_0}(H) cr_{t_0}(E \mid H) + c(H, E)}{cr_{t_0}(H) cr_{t_0}(E \mid H) + c(H, E) + cr_{t_0}(\neg H) cr_{t_0}(E \mid \neg H)}
\]

In: `DouvenPosterior[c_, pr_] := EvaluateProbability[Pr[H] \times Pr[E \mid H] + c + Pr[\neg H] \times Pr[E \mid \neg H], pr]`

For instance, if \(c(H, E) = \frac{1}{10}\), then we will obtain the value \(\frac{11}{16}\) for the (Douven) posterior.

In: `DouvenPosterior[1/10, prior]`

Out: \(\frac{11}{16}\)

Interestingly, there is no value of \(c(H, E) \in (0,1)\) such that the Bayesian \(t_2\) posterior \(cr_{t_2}(H \mid E \land A)\) is equal to the Douven posterior \(cr_{t_0}(H \mid E)\). Specifically, only if \(c(H, E) = 1\) will the two models will agree on the \(t_2\) posterior.

In: `Reduce[DouvenPosterior[c, prior] = 7/8, c]`

Out: \(c = 1\)

In: `DouvenPosterior[1, prior]`

Out: \(\frac{7}{8}\)

Ultimately, these two agents will “look the same” at \(t_2\) — over the \(\{H, E\}\) sub-algebra. To wit:

<table>
<thead>
<tr>
<th>(E)</th>
<th>(H)</th>
<th>(cr_{t_2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>(\frac{7}{8})</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>(\frac{1}{8})</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0</td>
</tr>
</tbody>
</table>

The advantage of the Bayesian approach is that — assuming probabilism — the prior \(cr_{t_0}\) completely determines what happens in subsequent updates (e.g., at \(cr_{t_1}\) and \(cr_{t_3}\)). There is no “extra parameter” \(c(H, E)\) that has to be fixed in order to determine \(cr_{t_2}\). Also, we don’t need any “exogenous perturbation” to the agent’s credences — it’s just standard conditionalization (first on \(E\) and then on \(A\)).

Moreover, I would suggest that the best way to fix the value of \(c(H, E)\) is by constructing a full Bayesian hypothetical prior and using it to “reverse engineer” a value for \(c(H, E)\) — as we did above, which led to the value \(c(H, E) = 1\). But, in this case, the reverse engineering leads to a value \([c(H, E) = 1]\), which is outside the range of values Douven assumes for \(c(H, E)\).

In fact, there seems to be no upper bound on \(c(H, E)\) that can be “reverse engineered” from Bayesian
theory. I have a function for creating models satisfying the above constraints, but which require arbitrarily large values of $c(H,E)$.

\[ \text{model}[] := \text{PrSAT} \{
\begin{align*}
\text{Pr}[H \mid E] &> \text{Pr}[H \mid \neg E], \\
\text{Pr}[H \mid E \land A] &> \text{Pr}[H \mid E \land \neg A], \\
\text{Pr}[A \mid E] &= \text{Pr}[A \mid \neg E], \\
\text{Pr}[H \mid A] &= \text{Pr}[H \mid \neg A], \\
\text{Pr}[H] &= \frac{1}{2}, \text{Pr}[E] = \frac{1}{2}, \text{Pr}[A] = \frac{1}{2}, \\
\text{Pr}[H \mid E \land A] &= \frac{\text{Pr}[H] \times \text{Pr}[E \mid H] + c}{\text{Pr}[H] \times \text{Pr}[E \mid H] + c + \text{Pr}[\neg H] \times \text{Pr}[E \mid \neg H]}
\end{align*}
\}, \text{Probabilities} \rightarrow \text{Regular, BypassSearch} \rightarrow \text{True}; \]

\[ \text{model}[100] \]

\[
\{ \begin{array}{l}
\{ a \rightarrow \{ a_2, a_5, a_6, a_8 \}, E \rightarrow \{ a_3, a_5, a_7, a_8 \}, H \rightarrow \{ a_4, a_6, a_7, a_8 \}, \\
\Omega \rightarrow \{ a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 \} \}, \{ a_1 \rightarrow \frac{10828}{162409}, a_2 \rightarrow \frac{40528}{162409}, a_3 \rightarrow \frac{119097}{649636}, \\
a_4 \rightarrow \frac{119097}{649636}, a_5 \rightarrow \frac{297}{649636}, a_6 \rightarrow \frac{297}{649636}, a_7 \rightarrow \frac{10828}{162409}, a_8 \rightarrow \frac{40528}{162409} \}
\end{array}\}
\]