1 The Observation Function

In order to incentivize our agent to take actions, we chose an observation function that is dependent on β, which in is dependent on the robot's actions. We call this observation the posterior observation function. This observation function essentially says "the human picks an observation proportional to the robot's belief in the desired object if the human had chosen that observation".

That is, we set

$$p(o|s) = p(o|\iota, \beta) \triangleq \eta p(\iota|o)_\beta \tag{1}$$

$$= \eta \frac{p(o|\iota)p(\iota)}{\sum_{\iota'} p(o|\iota')p(\iota')} \tag{2}$$

Note the distinction between $p(o|\iota, \beta)$ and $p(o|\iota)$. The former is the full observation function Ω, as we use in the POMDP. $p(o|\iota)$ is a base level observation function that is only dependent on the object. These are the unigram model for speech and the normally-distributed model for gesture described elsewhere.

Next, we will use we will set $p(\iota) = \beta(\iota)$. $p(\iota)$ describes the robot's belief in ι, or $b(\iota)$. However, since the human does not know this, it must use its estimate of b, β. Our new expression is

$$p(o|s) = \eta \frac{p(o|\iota)\beta(\iota)}{\sum_{\iota'} p(o|\iota')\beta(\iota')}$$

2 Toy Domain Example

In order to demonstrate and motivate this observation function, we will define a toy domain. In this toy domain, the states are the following: $\{AA, AB, BA, BB\}$ as well as β, which is a distribution over these states. The observations are $\{A\_, \_A, B\_, \_B\}$, which mean "the first character is an A", "the second character is an A", "the first character is B", "and the second character is a B" respectively. These observations are provided by the human. The agent can take actions CX which is informing the human that the agent believes the Xth character is a C, as well as "picking" the object or waiting. In our toy domain, for the base leve observation, the human always gives truthful observations, and has equal probability of generating an observation pertaining to a particular character. We will see how this affects the POMDP observation function, which incorporates β.

Consider the following situation:

Intially both b, the robot's belief about the human's desired object and β, the human's belief about the robot's belief, are uniform. The true state is AA.

t	b	β	a	o
0	[0.25, 0.25, 0.25, 0.25]	[0.25, 0.25, 0.25, 0.25]		

The robot then receives an observation $A\_$. The new beliefs are:

t	b	β	a	o
0	[0.25, 0.25, 0.25, 0.25]	[0.25, 0.25, 0.25, 0.25]		
1	[0.50, 0.50, 0, 0]	[0.25, 0.25, 0.25, 0.25]		$A\_$

Next, the robot can choose to take an action. If it chooses to wait, this is the resulting state

t	b	β	a	o
0	[0.25, 0.25, 0.25, 0.25]	[0.25, 0.25, 0.25, 0.25]		
1	[0.50, 0.50, 0, 0]	[0.25, 0.25, 0.25, 0.25]		$A\_$
2	[0.50, 0.50, 0, 0]	[0.25, 0.25, 0.25, 0.25]	wait	

Examine the probabilities of each observation.

$$\begin{aligned} p(A\_|AA, \beta) &= \frac{p(A\_)\beta(AA)}{p(A\_)\beta(AA) + p(A\_)\beta(AB) + p(A\_)\beta(BA) + p(A\_)\beta(BB)} \\ &= \frac{0.5 * 0.25}{0.5 * 0.25 + 0.5 * 0.25 + 0 + 0} \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} p(\_A|AA, \beta) &= \frac{p(\_A)\beta(AA)}{p(\_A)\beta(AA) + p(\_A)\beta(AB) + p(\_A)\beta(BA) + p(\_A)\beta(BB)} \\ &= \frac{0.5 * 0.25}{0.5 * 0.25 + 0 + 0.5 * 0.25 + 0} \\ &= 0.5 \end{aligned}$$

The probabilities of all other actions are 0, since we only give truthful observations.

Notice that this situation is unideal. The agent has equal probabilities of receiving either observation, even though the agent already knows that the first character is an A. Now, consider what would happen if the robot chose the action $A0$. We would get the following belief states:

t	b	β	a	o
0	[0.25, 0.25, 0.25, 0.25]	[0.25, 0.25, 0.25, 0.25]		
1	[0.50, 0.50, 0, 0]	[0.25, 0.25, 0.25, 0.25]		$A\_$
2	[0.50, 0.50, 0, 0]	[0.5, 0.5, 0, 0]	$A0$	

If we examine the probabilities again:

$$
\begin{aligned}
p(A\_|AA,\beta) &\propto \frac{p(A\_)\beta(AA)}{p(A\_)\beta(AA)+p(A\_)\beta(AB)+p(A\_)\beta(BA)+p(A\_)\beta(BB)} \\
&\propto \frac{0.5*0.5}{0.5*0.5+0.5*0.5+0+0} \\
&\propto 0.5 \\
&= \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
p(\_A|AA,\beta) &\propto \frac{p(\_A)\beta(AA)}{p(\_A)\beta(AA)+p(\_A)\beta(AB)+p(\_A)\beta(BA)+p(\_A)\beta(BB)} \\
&\propto \frac{0.5*0.5}{0.5*0.5+0+0+0} \\
&\propto 1 \\
&= \frac{2}{3}
\end{aligned}
$$

Now, we are more likely to get the observation $\_A$, which is more useful to us than the observation $A\_$, since it gives us the information we need to pick the correct object, AA.

3 Adaptation for Social Feedback Domain

In our domain, we have both speech and gesture, which we assume are conditionally independent given the state.

$$p(o|s) = p(l|s)p(g|s)$$

For each of language and gesture, we will use their own posterior observation function.

$$p(o|s) = \eta \frac{p(l|\iota)\beta(\iota)}{\sum_{\iota'} p(l|\iota')\beta(\iota')} \frac{p(g|\iota)\beta(\iota)}{\sum_{\iota'} p(g|\iota')\beta(\iota')}$$