Dynamic choice modeling

Even though I have written it many many times, the myriad of notations can still be confusing and call for some clarifications.

Reward function

$$u(s, a)$$

Total reward function

$$v(s, a, \epsilon) = u(s, a) + \epsilon_a$$

Continuation reward function

$$\tilde{u}(s, a) = u(s, a) + \beta E_{(s_{t+1}, \epsilon') | (a, s, \epsilon)} V(s_{t+1}, \epsilon')$$

Continuation total reward function

$$\tilde{v}(s, a, \epsilon) = \tilde{u}(s, a) + \epsilon_a$$

Value function

$$V(s, \epsilon) = \max_a \tilde{v}(s, a, \epsilon)$$

Expected value function

$$\bar{V}(s) = \int_\epsilon V(s, \epsilon) dF(\epsilon)$$

Conditional independence assumption

Thanks to the following assumption,

$$f((s_{t+1}, \epsilon') \mid (a, s, \epsilon)) = f(\epsilon') f(s_{t+1} \mid (a, s))$$

we have the relationship between value function and expected value function:

$$\bar{V}(s) = \int_\epsilon \left[\underbrace{\underbrace{\max_a u(s, a) + \beta E_{s_{t+1} | (a, s)} \bar{V}(s_{t+1})}_{\tilde{u}(s,a)} + \epsilon_a}_{\tilde{v}(s,a,\epsilon)} \right] dF(\epsilon)$$

which is used to sovle for the expected value function by solving for the fixed point of the following operator.

Dynamic choice estimation

Derive expression including CCP

First we need to rewrite the expected value function in terms of CCP

Approach 1

$$
\begin{aligned}
\bar{V}(s) &= \sum_a \Pr(a \mid s)\left[u(s,a) + E(\epsilon_a|s,a) + \beta E_{s_{t+1}|(a,s)} \bar{V}(s_{t+1})\right] \\
&= \sum_a \Pr(a \mid s)\left[\tilde{u}(s,a) + \gamma - \ln \Pr(a \mid s)\right]
\end{aligned}
$$

Because we have

$$
E(\epsilon_a|s,a) = \frac{1}{\Pr(a \mid s)} \int \epsilon_a 1\left\{\tilde{u}(s,a) + \epsilon_a > \tilde{u}(s,a') + \epsilon_{a'}\right\} dF(\epsilon)
$$

And in the case of T1EV,

$$
E(\epsilon_a|s,a) = \gamma - \ln \Pr(a \mid s)
$$

Approach 2

$$
\begin{aligned}
\bar{V}(s) &= \int \max_a \left[u(s,a) + \epsilon_a + \beta E_{s_{t+1}|(a,s)} \bar{V}(s_{t+1})\right] dF(\epsilon) \\
&= \gamma + \ln \sum_{a'} \exp\left(\tilde{u}(s,a')\right)
\end{aligned}
$$

At the same time, we have

$$
\Pr(a \mid s) = \frac{\exp(\tilde{u}(s,a))}{\sum_{a'} \exp(\tilde{u}(s,a'))}
$$

Then we can rewrite the expected value function in terms of CCP of a particular action a as follows:

$$
\begin{aligned}
\bar{V}(s) &= \gamma + \ln \exp(\tilde{u}(s,a)) - \ln \Pr(a \mid s) \\
&= \tilde{u}(s,a) - \ln \Pr(a \mid s) + \gamma
\end{aligned}
$$

Remark How to show the two approaches are equivalent?

Comparing

$$
\sum_a \Pr(a \mid s)\left[\tilde{u}(s,a) + \gamma - \ln \Pr(a \mid s)\right]
$$

with

$$
\tilde{u}(s,a) - \ln \Pr(a \mid s) + \gamma
$$

Notice that

$$
\ln \Pr(a \mid s) = \tilde{u}(s,a) - \ln \sum_{a'} \exp\left(\tilde{u}(s,a')\right)
$$

The proof is done.

Finite horizon case

Make use of terminal action

Time T In terminal action, we have that

$$\tilde{u}_T(s_T, a^*) = u_T(s_T, a^*)$$

Also, the CCP relationship is

$$\bar{V}_T(s_T) = u_T(s_T, a^*) - \ln \Pr(a^* \mid s_T) + \gamma$$

Time $T-1$ Then at $T-1$, we have

$$\tilde{u}_{T-1}(s_{T-1}, a) = u_{T-1}(s_{T-1}, a) + \beta E_{s_T|(a,s_{T-1})} \bar{V}_T(s_T)$$
$$= u_{T-1}(s_{T-1}, a) + \beta E_{s_T|(a,s_{T-1})} \left[u_T(s_T, a^*) - \ln \Pr(a^* \mid s_T) + \gamma\right]$$

Therefore, in time $T-1$, the model is essentially a static model with

$$\tilde{u}_{T-1}(s_{T-1}, a) = u_{T-1}(s_{T-1}, a) + \text{correction term}$$

Then the expected value function can be expressed as

$$\bar{V}_{T-1}(s_{T-1}) = \tilde{u}_{T-1}(s_{T-1}, a') - \ln \Pr_{T-1}(a' \mid s_{T-1}) + \gamma$$

Time $T-2$ Then at time $T-2$, we have

$$\tilde{u}_{T-2}(s_{T-2}, a) = u_{T-2}(s_{T-2}, a) + \beta E_{s_{T-1}|(a,s_{T-2})} \bar{V}_{T-1}(s_{T-1})$$

Similarly, we can roll backward.

Make use of renewal action at time $t+1$

Time t

$$\tilde{u}_t(s_t, a) = u_t(s_t, a) + \beta E_{s_{t+1}|(a,s)} \bar{V}_{t+1}(s_{t+1})$$

At time $t+1$, there is a renewal action a^*, then we can express

$$\bar{V}_{t+1}(s_{t+1}) = \tilde{u}_{t+1}(s_{t+1}, a^*) - \ln \Pr_{t+1}(a^* \mid s_{t+1}) + \gamma$$

Time $t+1$ Pick the renewal action a^*, we have

$$\tilde{u}_{t+1}(s_{t+1}, a^*) = u_{t+1}(s_{t+1}, a^*) + \beta E_{s_{t+2}|(a^*,s_{t+1})} \bar{V}_{t+2}(s_{t+2})$$

Renewal action means some kind of independence, let us write it out

$$\int (\tilde{u}_{t+1}(s_{t+1}, a^*) - \ln \Pr_{t+1}(a^* \mid s_{t+1}) + \gamma) dF_{s_{t+1}|(s,a)}(s_{t+1})$$

Plugging in $\tilde{u}_{t+1}(s_{t+1}, a^*)$ where

$$\tilde{u}_{t+1}(s_{t+1}, a^*) = u_{t+1}(s_{t+1}, a^*) + \beta E_{s_{t+2}|(a^*, s_{t+1})} \bar{V}_{t+2}(s_{t+2})$$

Therefore, we have a double integral

$$\int_{s_{t+1}} \int_{s_{t+2}} \bar{V}_{t+2}(s_{t+2}) dF_{s_{t+2}|(a^*, s_{t+1})}(s_{t+2}) dF_{s_{t+1}|(s,a)}(s_{t+1})$$

The idea of renewal action is that when we take action a^*, it doesn't matter what the state s_{t+1} is.

$$F_{s_{t+2}|(a^*, s_{t+1})}(s_{t+2}) = F_{s_{t+2}|(a^*)}(s_{t+2})$$

This is the idea.

Therefore, let's say at time t, take the difference between two actions a_1 and a_2, would give

$$\int_{s_{t+1}} \int_{s_{t+2}} \bar{V}_{t+2}(s_{t+2}) dF_{s_{t+2}|(a^*, s_{t+1})}(s_{t+2}) dF_{s_{t+1}|(s,a_1)}(s_{t+1})$$

$$- \int_{s_{t+1}} \int_{s_{t+2}} \bar{V}_{t+2}(s_{t+2}) dF_{s_{t+2}|(a^*, s_{t+1})}(s_{t+2}) dF_{s_{t+1}|(s,a_2)}(s_{t+1})$$

You couldn't combine the inner integral because $dF_{s_{t+1}|(s,a_1)}(s_{t+1})$ and $dF_{s_{t+1}|(s,a_2)}(s_{t+1})$ are different.

However, since

$$\int_{s_{t+2}} \bar{V}_{t+2}(s_{t+2}) dF_{s_{t+2}|(a^*, s_{t+1})}(s_{t+2}) = \int_{s_{t+2}} \bar{V}_{t+2}(s_{t+2}) dF_{s_{t+2}|(a^*)}(s_{t+2})$$

has nothing to do with s_{t+1}, we can take it out of the integral.

Then the difference becomes zero. Yeah.

Make use of renewal action at time $t + k$

Then the idea is that we can keep rolling forward until a period where there is a renewal action. For example, at time $t + k$, we can pick a renewal action. Then the integral term of $\bar{V}_{t+k+1}(s_{t+k+1})$ can be removed therefore we are free from estimating the expected value function.

And everything are expressed in terms of u and Pr without $\bar{V}$.