Temporal Difference Lambda

Given a sample path : $\{S_0, S_1, S_2, \dots S_T\}$, with state space $S_t \in \{s_1, s_2, s_3, \dots s_M\}$, where $t \in [0,T]$, the forward view and backward view of *temporal difference lambda* respectively are :

$$
\begin{array}{llll}
 & \Delta_{fwd} & = & \sum_{t=0}^{\infty} \alpha (G_{\lambda t}(s) - V(S_t)) 1_{S_t = s} & \textit{any limited path can be converted to } \infty \textit{ path by repeating last state} \\
 & \Delta_{back} & = & \sum_{t=0}^{\infty} \alpha \delta_t E_t(s) & \\
\textit{where} & G_t^{\lambda}(s) & = & (1-\lambda) \sum_{n=0} \lambda^n G_t^{(n)}(s) & \textit{where is (1-}\lambda\textit{) normalization factor as } \sum_{n=0} \lambda^n = (1-\lambda)^{-1} \\
 & G_t^{(n)}(s) & = & R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots + \gamma^{n-1} R_{t+n} + \gamma^n V(S_{t+n}) & \\
\textit{and} & \delta_t & = & R_{t+1} + \gamma V(S_{t+1}) - V(S_t) & \textit{which is one step look ahead error} \\
 & E_t(s) & = & (\lambda\gamma) E_{t-1}(s) + 1_{S_t = s} & \textit{called eligibility trace}
\end{array}
$$

Both views are equivalent. The forward view gives the intuition, while the backward view gives the implementation. Prior to proof, We have to unwire the recursive-form eligibility trace into iterative-form :

$$
\begin{array}{lll}
E_t(s) & = & (\lambda\gamma) E_{t-1}(s) + 1_{S_t = s} \qquad \textit{where } s \in \{s_1, s_2, s_3, \dots s_M\} \\
 & = & (\lambda\gamma)(\lambda\gamma E_{t-2}(s) + 1_{S_{t-1} = s}) + 1_{S_t = s} \\
 & = & (\lambda\gamma)^2 E_{t-2}(s) + (\lambda\gamma) 1_{S_{t-1} = s} + 1_{S_t = s} \\
 & = & (\lambda\gamma)^3 E_{t-3}(s) + (\lambda\gamma)^2 1_{S_{t-2} = s} + (\lambda\gamma) 1_{S_{t-1} = s} + 1_{S_t = s} \\
 & = & \dots \\
 & = & (\lambda\gamma)^t E_{t-t}(s) + \sum_{n=1}^{t} (\lambda\gamma)^{t-n} 1_{S_n = s} \\
 & = & \sum_{n=0}^{t} (\lambda\gamma)^{t-n} 1_{S_n = s} \qquad \textit{since } E_0(s) = 1_{S_0 = s}
\end{array}
$$

$$
\begin{array}{llll}
\text{i.e.} & E_0(s) & = & 1_{S_0 = s} \\
 & E_1(s) & = & 1_{S_1 = s} + (\lambda\gamma) 1_{S_0 = s} \\
 & E_2(s) & = & 1_{S_2 = s} + (\lambda\gamma) 1_{S_1 = s} + (\lambda\gamma)^2 1_{S_0 = s} \\
 & E_3(s) & = & 1_{S_3 = s} + (\lambda\gamma) 1_{S_2 = s} + (\lambda\gamma)^2 1_{S_1 = s} + (\lambda\gamma)^3 1_{S_0 = s}
\end{array}
$$

Lets start proving from forward view, there are 3 for loops, starting from outermost :

- backup starting from different states along one sample path
- sum of different TDs
- sum of different returns in one sample path

We have to group these 3 loops in forward view into 2 loops in backward view.

$$
\Delta_{fwd} = \begin{array}{l}
\alpha \left[(1-\lambda) \times \begin{pmatrix} + \lambda^0 (R_1 + \gamma V(S_1)) \\ + \lambda^1 (R_1 + \gamma R_2 + \gamma^2 V(S_2)) \\ + \lambda^2 (R_1 + \gamma R_2 + \gamma^2 R_3 + \gamma^3 V(S_3)) \\ + \lambda^3 (R_1 + \gamma R_2 + \gamma^2 R_3 + \gamma^3 R_4 + \gamma^4 V(S_4)) \\ + \dots \end{pmatrix} - V(S_0) \right] 1_{S_0 = s} + \\
\alpha \left[(1-\lambda) \times \begin{pmatrix} + \lambda^0 (R_2 + \gamma V(S_2)) \\ + \lambda^1 (R_2 + \gamma R_3 + \gamma^2 V(S_3)) \\ + \lambda^2 (R_2 + \gamma R_3 + \gamma^2 R_4 + \gamma^3 V(S_4)) \\ + \lambda^3 (R_2 + \gamma R_3 + \gamma^2 R_4 + \gamma^3 R_5 + \gamma^4 V(S_5)) \\ + \dots \end{pmatrix} - V(S_1) \right] 1_{S_1 = s} + \dots
\end{array}
$$

The target is to group all R_1 terms together, all R_2 terms together, and so on, making use of $\sum_{n=0}\lambda^n = (1-\lambda)^{-1}$. Thus removes 1 loop.

$$\Delta_{fwd} = \begin{array}{l} \alpha\left[(1-\lambda)\times\begin{pmatrix}(1-\lambda)^{-1}R_1+(1-\lambda)^{-1}(\lambda\gamma)R_2+(1-\lambda)^{-1}(\lambda\gamma)^2R_3+(1-\lambda)^{-1}(\lambda\gamma)^3R_4+\ldots \\ +\gamma[(\lambda\gamma)^0V(S_1)+(\lambda\gamma)^1V(S_2)+(\lambda\gamma)^2V(S_3)+(\lambda\gamma)^3V(S_4)+\ldots]\end{pmatrix}-V(S_0)\right]1_{S_0=s}+ \\ \alpha\left[(1-\lambda)\times\begin{pmatrix}(1-\lambda)^{-1}R_2+(1-\lambda)^{-1}(\lambda\gamma)R_3+(1-\lambda)^{-1}(\lambda\gamma)^2R_4+(1-\lambda)^{-1}(\lambda\gamma)^3R_5+\ldots \\ +\gamma[(\lambda\gamma)^0V(S_2)+(\lambda\gamma)^1V(S_3)+(\lambda\gamma)^2V(S_4)+(\lambda\gamma)^3V(S_5)+\ldots]\end{pmatrix}-V(S_1)\right]1_{S_1=s}+\ldots \end{array}$$

$$= \begin{array}{l} \alpha\left[\begin{pmatrix}R_1+(\lambda\gamma)R_2+(\lambda\gamma)^2R_3+(\lambda\gamma)^3R_4+\ldots \\ +\gamma[(\lambda\gamma)^0V(S_1)+(\lambda\gamma)^1V(S_2)+(\lambda\gamma)^2V(S_3)+(\lambda\gamma)^3V(S_4)+\ldots] \\ -[(\lambda\gamma)^1V(S_1)+(\lambda\gamma)^2V(S_2)+(\lambda\gamma)^3V(S_3)+(\lambda\gamma)^4V(S_4)+\ldots]\end{pmatrix}-V(S_0)\right]1_{S_0=s}+ \\ \alpha\left[\begin{pmatrix}R_2+(\lambda\gamma)R_3+(\lambda\gamma)^2R_4+(\lambda\gamma)^3R_5+\ldots \\ +\gamma[(\lambda\gamma)^0V(S_2)+(\lambda\gamma)^1V(S_3)+(\lambda\gamma)^2V(S_4)+(\lambda\gamma)^3V(S_5)+\ldots] \\ -[(\lambda\gamma)^1V(S_2)+(\lambda\gamma)^2V(S_3)+(\lambda\gamma)^3V(S_4)+(\lambda\gamma)^4V(S_5)+\ldots]\end{pmatrix}-V(S_1)\right]1_{S_1=s}+\ldots \end{array}$$

Then group all terms according to $(\lambda\gamma)$, $(\lambda\gamma)^2$, $(\lambda\gamma)^3$ and so on.

$$\Delta_{fwd} = \alpha\begin{bmatrix}+(\lambda\gamma)^0(R_1+\gamma V(S_1)-V(S_0)) \\ +(\lambda\gamma)^1(R_2+\gamma V(S_2)-V(S_1)) \\ +(\lambda\gamma)^2(R_3+\gamma V(S_3)-V(S_2)) \\ +(\lambda\gamma)^3(R_4+\gamma V(S_4)-V(S_3)) \\ +\ldots\end{bmatrix}1_{S_0=s}+\alpha\begin{bmatrix}+(\lambda\gamma)^0(R_2+\gamma V(S_2)-V(S_1)) \\ +(\lambda\gamma)^1(R_3+\gamma V(S_3)-V(S_2)) \\ +(\lambda\gamma)^2(R_4+\gamma V(S_4)-V(S_3)) \\ +(\lambda\gamma)^3(R_5+\gamma V(S_5)-V(S_4)) \\ +\ldots\end{bmatrix}1_{S_1=s}+\ldots$$

$$= \begin{bmatrix}+\alpha(R_1+\gamma V(S_1)-V(S_0))\times(1_{S_0=s}) \\ +\alpha(R_2+\gamma V(S_2)-V(S_1))\times(1_{S_1=s}+(\lambda\gamma)1_{S_0=s}) \\ +\alpha(R_3+\gamma V(S_3)-V(S_2))\times(1_{S_2=s}+(\lambda\gamma)1_{S_1=s}+(\lambda\gamma)^21_{S_0=s}) \\ +\alpha(R_4+\gamma V(S_4)-V(S_3))\times(1_{S_3=s}+(\lambda\gamma)1_{S_2=s}+(\lambda\gamma)^21_{S_1=s})+(\lambda\gamma)^31_{S_0=s}) \\ +\ldots\end{bmatrix}$$

$$= \sum_{t=0}^{\infty}\alpha(R_{t+1}+\gamma V(S_{t+1})-V(S_t))E_t(s)$$

$$= \sum_{t=0}^{\infty}\alpha\delta_tE_t(s) \qquad where\ \ \delta_t = R_{t+1}+\gamma V(S_{t+1})-V(S_t)$$

$$= \Delta_{back}$$

Therefore, it is equivalent to backward view temporal difference lambda.