6-3

The change of $V(A)$ implies that the first episode terminates on the left. I'll explain this reason. First, transitions between states **except terminal states** cause no change of $V(S)$. Thus, we can focus on only transitions to terminal states in the episode. Transitions to the **right** terminal state affect not $V(A)$ but $V(E)$. Therefore, transitions to the **left** one lower $V(A)$. Concretely,

$$\begin{aligned} V'(A) &\leftarrow V(A) - \alpha\,(R + V(L) - V(A)) \\ &= 0.5 + 0.1 \cdot (0 + 0 - 0.5) \\ &= 0.45 \end{aligned}$$

6-6

- DP Approach
  Since this problem gives environment's dynamics and need no policy, we can use DP approach. DP approach calculate acculate value functions by value iteration. (but need some calculation cost)

- Mathematical Approach
  We can also use probablistic math. More specifically, we use the fact that $V(S)$ corresponds to probability of terminating on the Right from state $S.(\doteq P_S(R))$
  For example, Let's calculate $V(E)$.

$$\begin{aligned} V(E) = P_E(R) &= 1 - P_E(L) \\ &= 1 - P_E(D) \cdot P_D(L) \\ &= 1 - P_E(D) \cdot [P_D(C) \cdot P_C(L) + P_D(E) \cdot P_E(L)] \end{aligned}$$

Then,

$$\begin{aligned} 1 - P_E(L) &= 1 - P_E(D) \cdot [P_D(C) \cdot P_C(L) + P_D(E) \cdot P_E(L)] \\ P_E(L) &= P_E(D) \cdot [P_D(C) \cdot P_C(L) + P_D(E) \cdot P_E(L)] \\ P_E(L) &= 0.5 \cdot [0.5 \cdot 0.5 + 0.5 \cdot P_E(L)] \;\; (\because P_C(L) = 0.5) \\ 6P_E(L) &= 1 \\ P_E(L) &= 1/6 \end{aligned}$$

Therefore,

$$V(E) = 5/6$$

Likewise, we can calculate other state value.