(۳

Monte Carlo:

$$V(S_t) = V(S_t) + \alpha \left[G_t - V(S_t) \right]$$

$$G_t = R_{t+۱} + \gamma R_{t+۲} + \gamma^{۲} R_{t+۳} + \dots$$

$\alpha = ۰/۵ \, , \, \gamma = ۱$

First_visit:

$$V(S_۱) = V(S_۱) + ۰/۵ \left(G_۱ - V(S_۱) \right)$$

$$G_۱ = R_۲ + \gamma R_۳ + \gamma^{۲} R_۸ + \dots + \gamma^{۱۳} R_{۲۳}$$

$$= ۰ + ۰ + ۰ + ۰ + ۰ + ۰ + ۰ - ۵ + ۰ - ۵ + ۰ + ۰ + ۰ + ۰ + ۱۰$$

$$= ۰ \Rightarrow V(S_۱) = ۰ + ۰/۵ (۰ - ۰) = ۰$$

$$V(S_۲) = V(S_۲) + ۰/۵ \left(G_۲ - V(S_۲) \right)$$

$$G_۲ = R_۳ + \gamma R_۸ + \gamma^{۲} R_۷ + \dots + \gamma^{۱۲} R_{۲۳}$$

$$= ۰ + ۰ + ۰ + ۰ + ۰ + ۰ - ۵ + ۰ - ۵ + ۰ + ۰ + ۰ + ۰ + ۱۰$$

$$= ۰ \Rightarrow V(S_۲) = ۰ + ۰/۵ (۰ - ۰) = ۰$$

$V(S_{۳}) = ۰ + ۰٫۵ (\overset{۰}{\cancel{G_{۳}}} - ۰) = ۰$

$V(S_{۸}) = ۰ + ۰٫۵ (\overset{۰}{\cancel{G_{۸}}} - ۰) = ۰$

$V(S_{۷}) = ۰ + ۰٫۵ (\overset{۰}{\cancel{G_{۷}}} - ۰) = ۰$

$V(S_{۱۲}) = ۰ + ۰٫۵ (\overset{۰}{\cancel{G_{۱۲}}} - ۰) = ۰$

$V(S_{۱۴}) = ۰ + ۰٫۵ (\overset{۰}{\cancel{G_{۱۴}}} - ۰) = ۰$

$V(S_{۲۰}) = ۰ + ۰٫۵ (\overset{۰}{\cancel{G_{۲۰}}} - ۰) = ۰$

$V(S_{۲۱}) = ۰ + ۰٫۵ (G_{۲۱} - ۰)$

$G_{۲۱} = R_{۲۲} + \gamma R_{۲۱} + \gamma^{۲} R_{۲۲} + \gamma^{۳} R_{۱۷} + \gamma^{۴} R_{۱۸} + \gamma^{۵} R_{۲۳}$

$= ۰ - ۵ + ۰ + ۰ + ۰ + ۱۰ = ۵$

$\Rightarrow V(S_{۲۱}) = ۰ + ۰٫۵ (۵ - ۰) = ۲٫۵$

$V(S_{۲۲}) = ۰ + ۰٫۵ (\overset{۵}{\cancel{G_{۲۲}}} - ۰) = ۲٫۵$

$V(S_{۱۷}) = ۰ + ۰٫۵ (G_{۱۷} - ۰)$

$G_{۱۷} = R_{۱۸} + \gamma R_{۲۳} = ۱۰ \Rightarrow V(S_{۱۷}) = ۵$

$V(S_{۱۸}) = ۰ + ۰/۵(\cancel{G_{۱۸}} - ۰) = ۵$

$V(S_{۲۳}) \Longrightarrow$ terminal state $\Longrightarrow V(S_{۲۳}) = ۰$

every-visit :

$V(S_{۱}) = V(S_{۱}) + ۰/۵(G_{۱} - V(S_{۱})$

برای G_t : هر بار که S_t اتفاق بیفتد یک جمع جدید از پاداش ها داریم

و در نهایت میانگین میگیریم.

$V(S_{۱}) = V(S_{۱}) + ۰/۵(G_{۱} - V(S_{۱}))$

$G_{۱}$: $S_{۱}$ که در تمام اپیزود یکبار اتفاق افتاده است پس یک عبارت جمع داریم

$$\Rightarrow G_{۱} = ۰+۰+۰+۰+۰ \\ +۰+۰-۵+۰-۵ \\ +۰+۰+۱۰+۱۰ \\ = ۰$$

$\Rightarrow V(S_{۱}) = ۰ + ۰/۵(۰ - ۰) = ۰$

$V(S_{۲}) = ۰ + ۰/۵(\underset{۰}{\cancel{G_{۲}}} - ۰) = ۰$

$V(S_{۳}) = ۰ + ۰/۵(\underset{۰}{\cancel{G_{۳}}} - ۰) = ۰$

$$V(S_{۸}) = ۰ + ۰٫۵\left(\underset{۰}{\cancel{G_{۸}}} - ۰\right) = ۰$$

$$V(S_{۱۲}) = ۰$$

$$V(S_{۱۴}) = ۰$$

$$V(S_{۲۰}) = ۰$$

$$V(S_{۲۱}) = ۰ + ۰٫۵\left(G_{۲۱} - ۰\right)$$

$$G_{۲۱} = \Big[\left(R_{۲۲} + \gamma R_{۲۱} + \gamma^{۲} R_{۲۲} + \gamma^{۳} R_{۱۷} + \gamma^{۴} R_{۱۸} + \gamma^{۵} R_{۲۳}\right)$$

$$+ \left(R_{۲۲} + \gamma R_{۱۷} + \gamma^{۲} R_{۱۸} + \gamma^{۳} R_{۲۳}\right)\Big] / ۲$$

$$= \Big[\left(۰ - ۵ + ۰ + ۰ + ۰ + ۱۰\right) + \left(۰ + ۰ + ۰ + ۱۰\right)\Big] / ۲$$

$$= \frac{۵ + ۱۰}{۲} = ۷٫۵ \Rightarrow V(S_{۲۱}) = ۳٫۷۵$$

$$V(S_{۲۲}) = ۰ + ۰٫۵\left(G_{۲۲} - ۰\right)$$

$$G_{۲۲} = \Big[\left(R_{۲۱} + \gamma R_{۲۲} + \gamma^{۲} R_{۱۷} + \gamma^{۳} R_{۱۸} + \gamma^{۴} R_{۲۳}\right) +$$

$$\left(R_{۱۷} + \gamma R_{۱۸} + \gamma R_{۲۳}\right)\Big] / ۲ = \Big[\left(-۵ + ۰ + ۰ + ۰ + ۱۰\right)$$

$$+ \left(۰ + ۰ + ۱۰\right)\Big] / ۲ = ۷٫۵ \Rightarrow V(S_{۲۲}) = ۳٫۷۵$$

$$V(S_{۱۷}) = ۰ + ۰٫۵\,(G_{۱۷} - V(S_{۱۷}))$$

$$G_{۱۷} = (R_{۱۸} + \gamma R_{۲۳}) = ۰ + ۱۰ = ۱۰ \Rightarrow V(S_{۱۷}) =$$

$$۰ + ۰٫۵\,(۱۰ - ۰) = ۵$$

$$V(S_{۱۸}) = ۰ + ۰٫۵\,(\overset{۱۰}{\cancel{G}}_{۱۸} - V(S_{۱۸})) = ۵$$

$$V(S_{۲۳}) \Rightarrow \text{terminal\_state} \Rightarrow V(S_{۲۳}) = ۰$$