Read Chapter 5 and my paper using this RW

A RW with 3 states

States 0, 1, 2

$$P = \begin{array}{ccc} 0 & 1 & 2 \\ q\_0 & 1\text{-}q\_0 & 0 \\ 1\text{-}q\_1/2 & q\_1 & 1\text{-}q\_1/2 \\ 0 & 1\text{-}q\_2 & q\_2 \end{array}$$

where (q_0=1/3, q_1=1/2, q_2=1/3)

m_0=1+ (1-q_0) *m_1. = (1+0)*q_0+ (1+m_1)**(1-q_0)

m_1=1+0*(1-q_1/2) +q_1*m_1+(1-q_1/2) *m_2

m_2=1 +(1-q_2)*m_1+q_2*m_2

For computing w_0, w_1, w_2, g, s_0,s_1,s_2 take

r_0=10,
r_1=20,
r_3=100.

β_1,β_2, … cycles of returns to a `ground state' 0.

For the MDP version

I)
Introduce 2 actions in state 1 (in some states).

action a11 as above (ie, p_{1a_11}=(1-q_1/2, q_1,1-q_1/2) and r_{1a_11}=20)

action a12 with p_{1a_12}=(1-q_1/3,q_1,1-2*q_1/3) and r_{1a_12}=10)

We have 2 policies, depending on the action in state 1.

II) Introduce 2 actions in state 0,1,2

action a01 as above (ie, p_{1a_01}, and r_{1a_11} as above)
action a02 with p_{1a_12}=(q_0, 1-q_0/2,1-q_0/2) and r_{1a_12}=1)

action a11 as above (ie, p_{1a_11}=(1-q_1/2, q_1,1-q_1/2) and r_{1a_11}=20)
action a12 with p_{1a_12}=(1-q_1/3, q_1,1-q_2/2) and r_{1a_12}=1)

action a21 as above (ie, p_{2a_21}, and r_{2a_21} as above)
action a22 with p_{2a_21}=1-q_2/3, 1-2*q_2/3,q_2) and r_{2a_22}=50)

We have 8=2^3 policies, depending on the actison in states 0, 1, 2.

Idea is to find optimal policies using

$$\bar{g}_n - U_n^g \leq g \leq \bar{g}_n - L_n^g,$$