The math derivation for Reinforcement Learning Assignment 1

At each of discrete time step: $t = 0, 1, \cdots, T-1$, we need to allocate W_t. We assume $T = 10$. And we allocate x_t on the risky asset and $W_t - x_t$ on riskless asset.

The return rate of risky asset follow the distribution,

$$P(Y_t = a) = p, \quad P(Y_t = b) = 1 - p$$

Set the discount factor is γ. We want to maximize the utility of wealth at $t = T$,

$$U(W_T) = \frac{1 - e^{-kW_T}}{k} \quad (k \neq 0)$$

Our goal is,

$$\max \mathbb{E}\left[\gamma^{T-t} \cdot \frac{1 - e^{-kW_T}}{k} \mid (t, W_t)\right]$$

It is equivalent to,

$$\max \mathbb{E}\left[\frac{-e^{-kW_T}}{k} \mid (t, W_t)\right]$$

We consider $\pi_t(W_t) = x_t$, and the optimal policy is $\pi_t^*(W_t) = x_t^*$.

It is easy to get the relation that,

$$W_{t+1} = x_t \cdot (1 + Y_t) + (W_t - x_t) \cdot (1 + r) = x_t \cdot (Y_t - r) + W_t \cdot (1 + r)$$

Then we can define the value function,

$$V_t^\pi(W_t) = \mathbb{E}_\pi\left[\frac{-e^{-kW_T}}{k} \mid (t, W_t)\right]$$

Then the optimal value function at time t ($\forall t = 0, 1, \ldots, T-1$) as:

$$V_t^*(W_t) = \max_\pi V_t^\pi(W_t) = \max_\pi \left\{\mathbb{E}_\pi\left[\frac{-e^{-kW_T}}{k} \mid (t, W_t)\right]\right\}$$

The Bellman Optimality Equation is,

$$V_t^*(W_t) = \max_{x_t} Q_t^*(W_t, x_t) = \max_{x_t} \left\{\mathbb{E}_{Y_t}\left[V_{t+1}^*(W_{t+1})\right]\right\}$$

$\forall t = 0, 1, \ldots, T-2$, and

$$V_{T-1}^*(W_{T-1}) = \max_{x_{T-1}} Q_{T-1}^*(W_{T-1}, x_{T-1}) = \max_{x_{T-1}} \left\{\mathbb{E}_{Y_{T-1}}\left[\frac{-e^{-kW_T}}{k}\right]\right\}$$

where Q_t^* is the Optimal Action-Value Function at time t, $\forall t = 0, 1, \ldots, T-1$.
We make a guess for the functional form of the Optimal Value Function as,

$$V_t^*(W_t) = -b_t \cdot e^{-c_t \cdot W_t}$$

where b_t, c_t are independent of the wealth W_t for all $t = 0, 1, \ldots, T-1$.

Then we can express the BOE,

$$\begin{aligned} V_t^*\left(W_t\right) &= \max_{x_t}\left\{\mathbb{E}_{Y_t}\left[-b_{t+1}\cdot e^{-c_{t+1}\cdot W_{t+1}}\right]\right\} \\ &= \max_{x_t}\left\{\mathbb{E}_{Y_t}\left[-b_{t+1}\cdot e^{-c_{t+1}\cdot\left(x_t\cdot\left(Y_t-r\right)+W_t\cdot(1+r)\right)}\right]\right\} \end{aligned}$$

With the distribution of Y_t, we can get that,

$$V_t^*\left(W_t\right)=\max_{x_t}\left\{-pb_{t+1}e^{-c_{t+1}\left[x_t(a-r)+W_t(1+r)\right]}-(1-p)b_{t+1}e^{-c_{t+1}\left[x_t(b-r)+W_t(1+r)\right]}\right\}$$

And we know that, $V_t^*\left(W_t\right)=\max_{x_t} Q_t^*\left(W_t,x_t\right)$. Hence we have,

$$Q_t^*\left(W_t,x_t\right)=-pb_{t+1}e^{-c_{t+1}\left[x_t(a-r)+W_t(1+r)\right]}-(1-p)b_{t+1}e^{-c_{t+1}\left[x_t(b-r)+W_t(1+r)\right]}$$

We take the derivation,

$$\frac{\partial Q_t}{\partial x_t}=0$$

We can get that, for the optimal x_t^*, it satisties,

$$e^{c_{t+1}(b-a)x_t^*}=\frac{(b-r)\,(p-1)}{(a-r)p}$$

Hence, the optimal action is,

$$x_t^*=\frac{1}{c_{t+1}\,(b-a)}\ln\left[\frac{(b-r)\,(1-p)}{(r-a)p}\right]$$

Put this result into the formula of $V_t^*\left(W_t\right)$, we get that,

$$V_t^*(W_t)_{x_t=x_t^*}=-Ab_{t+1}e^{c_{t+1}W_t(1+r)}$$

where A is,

$$A=p\left[\frac{(b-r)\,(1-p)}{(r-a)p}\right]^{-\frac{a-r}{b-a}}+(1-p)\left[\frac{(b-r)\,(1-p)}{(r-a)p}\right]^{-\frac{b-r}{b-a}}$$

Compare to the form,

$$V_t^*\left(W_t\right)=-b_t\cdot e^{-c_t\cdot W_t}$$

We can get that,

$$b_t=b_{t+1}A,\quad c_t=c_{t+1}\,(1+r)$$

To solve b_t and c_t, we consider that,

$$V_{T-1}^*\left(W_{T-1}\right)=\max_{x_{T-1}}\left\{\mathbb{E}_{Y_{T-1}}\left[\frac{-e^{-k\left(x_{T-1}\cdot\left(Y_{T-1}-r\right)+W_{T-1}\cdot(1+r)\right)}}{k}\right]\right\}$$

Solving the $\partial V_{T-1}^*\left(W_{T-1}\right)/\partial x_{T-1}=0$, we get the optimal x_{T+1} is,

$$x_{T-1}^{*}=\frac{1}{k\left(b-a\right)}\ln\left[\frac{\left(b-r\right)\left(1-p\right)}{(r-a)p}\right]$$

Then $V_{T-1}^{*}\left(W_{T-1}\right)$ is,

$$\begin{aligned}V_{T-1}^{*}\left(W_{T-1}\right)&=\left\{\frac{-p}{k}\left[\frac{\left(b-r\right)\left(1-p\right)}{(r-a)p}\right]^{-\frac{a-r}{b-a}}+\frac{-\left(1-p\right)}{k}\left[\frac{\left(b-r\right)\left(1-p\right)}{(r-a)p}\right]^{-\frac{b-r}{b-a}}\right\}e^{-k(1+r)W_{T}}\\&=\frac{-A}{k}e^{-k(1+r)W_{T}}\end{aligned}$$

Compare to the form,

$$V_{T-1}^{*}\left(W_{T-1}\right)=-b_{T-1}\cdot e^{-c_{T-1}\cdot W_{T-1}}$$

We can get that,

$$b_{T-1}=\frac{-A}{k},\quad c_{T-1}=k\left(1+r\right)$$

Use the equations that,

$$b_{t}=b_{t+1}A,\quad c_{t}=c_{t+1}\left(1+r\right)$$

We can get that,

$$b_{t}=b_{T-1}A^{T-t-1}=\frac{-1}{k}A^{T-t}$$

$$c_{t}=k(1+r)^{T-t}$$

where,

$$A=p\left[\frac{\left(b-r\right)\left(1-p\right)}{(r-a)p}\right]^{-\frac{a-r}{b-a}}+\left(1-p\right)\left[\frac{\left(b-r\right)\left(1-p\right)}{(r-a)p}\right]^{-\frac{b-r}{b-a}}$$

So the optimal policy,

$$\pi_{t}^{*}(W_{t})=x_{t}^{*}=\frac{1}{k(1+r)^{T-t-1}\left(b-a\right)}\ln\left[\frac{\left(b-r\right)\left(1-p\right)}{(r-a)p}\right]$$

The Optimal Value Function is,

$$V_{t}^{*}\left(W_{t}\right)=-Ab_{t+1}e^{c_{t+1}W_{t}(1+r)}=\frac{1}{k}A^{T-t}e^{k(1+r)^{T-t-1}W_{t}}$$

We can also get the Optimal Action-Value Function. [The result is too long so it is not shown here.]