1. Derive the solution to Merton's Portfolio problem for the case of the log($\cdot$) Utility function. Note that the derivation in the textbook is for CRRA Utility function with $\gamma \neq 1$ and the case of the log($\cdot$) Utility function was left as an exercise to the reader.

   Therefore, the infinitesimal change in wealth dW_t from time t to time $t + dt$ is given by:

   $$dW_t = ((r + \pi_t \cdot (\mu - r)) \cdot W_t - c_t) \cdot dt + \pi_t \cdot \sigma \cdot W_t \cdot dz_t \qquad (6.1)$$

   Note that this is an Ito process defining the stochastic evolution of wealth.
   Our goal is to determine optimal $(\pi(t, W_t), c(t, W_t))$ at any time t to maximize:

   $$\mathbb{E}[\int_t^T \frac{e^{-\rho(s-t)} \cdot c_s^{1-\gamma}}{1-\gamma} \cdot ds + \frac{e^{-\rho(T-t)} \cdot B(T) \cdot W_T^{1-\gamma}}{1-\gamma} \mid W_t]$$

   maximize $\quad E[\int_t^T e^{-\rho(S-t)} \log(c_s) \, ds + e^{-\rho(T-t)} \, \xi \, \log(W_T) \mid W_t]$

   $$\rho \cdot V^*(t, s_t) \cdot dt = \max_{a_t \in \mathcal{A}_t}\{\mathbb{E}_{(t,s_t,a_t)}[dV^*(t, s_t)] + \mathcal{R}(t, s_t, a_t) \cdot dt\}$$

   $$\rho \cdot V^*(t, \boldsymbol{s}_t) = \max_{a_t \in \mathcal{A}_t}\{\frac{\partial V^*}{\partial t} + (\nabla_{\boldsymbol{s}}V^*)^T \cdot \boldsymbol{\mu}_t + \frac{1}{2}Tr[\boldsymbol{\sigma}_t^T \cdot (\Delta_{\boldsymbol{s}}V^*) \cdot \boldsymbol{\sigma}_t] + \mathcal{R}(t, \boldsymbol{s}_t, a_t)\}$$

   $$\max_{\pi_t, c_t}\{\mathbb{E}_t[dV^*(t, W_t) + \frac{c_t^{1-\gamma}}{1-\gamma} \cdot dt\} = \rho \cdot V^*(t, W_t) \cdot dt$$

   $$\max_{\pi_t, c_t} \{ E[dV^*(t, W_t)] + \log(c_t) \, dt \} = \rho \, V^*(t, W_t) \, dt$$

   $$\max_{\pi_t, c_t}\{\frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial W_t} + \frac{\partial^2 V^*}{\partial W_t^2} \cdot \frac{\pi_t^2 \cdot \sigma^2 \cdot W_t^2}{2} + \frac{c_t^{1-\gamma}}{1-\gamma}\} = \rho \cdot V^*(t, W_t) \qquad V^*(T, W_T) = \epsilon^\gamma \cdot \frac{W_T^{1-\gamma}}{1-\gamma}$$

   $= \xi \log(W_T)$

   $$\max_{\pi_t, c_t} \{ \underbrace{\frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial W_t} + \frac{\partial^2 V^*}{\partial W_t^2} \cdot \frac{\pi_t^2 \, \sigma^2 W_t^2}{2} + \log(c_t)}_{\Phi(t, W_t, \pi_t, c_t)} \} = \rho \cdot V^*(t, W_t)$$

   $\frac{\partial \Phi}{\partial \pi_t} = \qquad (\mu - r) \cdot \frac{\partial V^*}{\partial W_t} + \frac{\partial^2 V^*}{\partial W_t^2} \cdot \pi_t \cdot \sigma^2 \cdot W_t = 0$

   $$\Rightarrow \pi_t^* = \frac{-\frac{\partial V^*}{\partial W_t} \cdot (\mu - r)}{\frac{\partial^2 V^*}{\partial W_t^2} \cdot \sigma^2 \cdot W_t}$$

   $\frac{\partial \Phi}{\partial c_t} = \qquad -\frac{\partial V^*}{\partial W_t} + (c_t^*)^{-\gamma} = 0 \qquad\qquad -\frac{\partial V^*}{\partial W_t} + \frac{1}{c_t^*} = 0$

   $$\Rightarrow c_t^* = (\frac{\partial V^*}{\partial W_t})^{\frac{-1}{\gamma}} \qquad\qquad \Rightarrow c_t^* = \frac{\partial W_t}{\partial V^*}$$

$$\frac{\partial V^*}{\partial t} - \frac{(\mu - r)^2}{2\sigma^2} \cdot \frac{(\frac{\partial V^*}{\partial W_t})^2}{\frac{\partial^2 V^*}{\partial W_t^2}} + \frac{\partial V^*}{\partial W_t} \cdot r \cdot W_t + \frac{\gamma}{1 - \gamma} \cdot \left(\frac{\partial V^*}{\partial W_t}\right)^{\frac{\gamma - 1}{\gamma}} = \rho \cdot V^*(t, W_t)$$

Stanford CME 241 (Winter 2021) - Assignment 7

Assignments:

1. Derive the solution to Merton's Portfolio problem for the case of the $\log(\cdot)$ Utility function. Note that the derivation in the textbook is for CRRA Utility function with $\gamma \neq 1$ and the case of the $\log(\cdot)$ Utility function was left as an exercise to the reader.

2. One of the reasons the backward induction solution in rl/chapter7/asset_alloc_discrete.py is slow is that we work with a generic Distribution type for risky_return_distributions , which means we have to sequentially sample from it to create the states distribution (in method get_states_distribution) that can be passed as input to back_opt_qvf. Modify the code to create a special type of distribution for the returns of the risky asset so we have a direct way of obtaining the probability distribution of the risky asset price at any time step (and hence, the probability distribution of wealth at any time step). With a direct way to obtain probability distribution of states at any time step, we can speed up the code considerably.

3. Sketch out a rough design for the following MDP:

   - Your job is such that you can spend a fraction α of each day working and the remaining fraction $1 - \alpha$ of your day learning.
   - Each minute that you spend in a day working earns you at the rate of $f(s)$ dollars per minutes where s is your current skill level.
   - Each minute you spend on learning improves your skill level by a certain factor $g(s)$ where s is your current skill level.
   - You could lose your job any day with a probability p.
   - While unemployed, you do not have access to learning, so your skill level decays exponentially with half life λ.
   - If you've lost your job, you could be offered the job back with a probability $h(s)$ where s is your current skill level.

   What should be your optimal policy for fraction of time on learning on any given day, if your goal is to maximize your Expected (Discounted) Lifetime Utility of Earnings? Think about variants such as finite-horizon and infinite-horizon. What if there were multiple skills or multiple job options? You can also factor in the fact that you need to consume your earnings, and so consumption quantity on any given day will also be an action, and then the objective would be Expected (Discounted) Lifetime Utility of Consumption.

   This question is open-ended and is meant as a thought exercise. But do try to sketch out the states, actions, rewards and transition function.