Problem 3 of Assignment 7

1. At first thought, state can be of two variables: 1) whether I am employed (1) or not (0) today, and 2) my skill level, s. My action every day is a continuous variable α. If I am in state (0,s), I can only choose $\alpha = 1$ since I do not have access to learning if unemployed. I can be transformed into state (0,$s \cdot e^{-\lambda}$) with a probability of $1 - h(s)$ and into state (1,s) with a probability $h(s)$. My rewards for state (0,s) is 0. If I am in state (1,s), I can be transformed into state (1,$s(1 + (1 - \alpha)g(s))$) with a probability of $1 - p$ and into state (0,$s(1 + (1 - \alpha)g(s))$) with a probability of p. My rewards for state (1,s) is $\alpha f(s)$. Thus, in summary, we have the following:

   - Mathematically, states are: $\mathcal{S} = \{(1, s), (0, s) | \forall s \geq 0 \in \mathbb{R}\}$. The first variable is an indicator representing whether I am employed today, and the second variable is a non-negative variable representing my skill level.
   - Whether it is finite or infinite horizon does change the formulation. If it is finite horizon, we can add the day t into the state formulation to account for non-terminal and terminal states. That is, we can have non-terminal states $\mathcal{N} = \{(1, s, t), (0, s, t) | \forall s \geq 0 \in \mathbb{R} for t \in \{0, 1, ..., T - 1\}\}$, and terminal states $\mathcal{N} = \{(1, s, T), (0, s, T) | \forall s \geq 0 \in \mathbb{R}\}$, with T denoting the terminal time. If it is infinite horizon, we can use the original formulation. Here we assume infinite horizon to continue.
   - Actions are: $\mathcal{A} = \alpha \in [0, 1]$.
   - Rewards are, in this problem, every day paychecks.
   - Transition probabilities in this problem are:
     - $\mathcal{P}_R(x, a, r, x') = p$, if $x = (1, s)$, $x' = (0, s(1 + (1 - \alpha)g(s)))$, $r = \alpha f(s)$.
     - $\mathcal{P}_R(x, a, r, x') = 1 - p$, if $x = (1, s)$, $x' = (1, s(1 + (1 - \alpha)g(s)))$, $r = \alpha f(s)$.
     - $\mathcal{P}_R(x, a, r, x') = h(s)$, if $x = (0, s)$, $x' = (1, s)$, $r = 0$.
     - $\mathcal{P}_R(x, a, r, x') = 1 - h(s)$, if $x = (0, s)$, $x' = (0, s \cdot e^{-\lambda})$, $r = 0$.
   - Then, to maximize my Expected (Discounted) Lifetime Utility of Earnings, I need to model my Utility consumption function first. When there were multiple skills or jobs to choose from, the state formulation need to include my skill levels for all the skills, and use a categorical variable instead of an indicator variable to model jobs (they may have different levels of payments). I may also need to pick a discount factor, which make some sense to be just the risk-free rate. If consumption quantity on any given day become a part of the action, say c, we will also need to consider the return and risk of the remaining $1 - c$ we put into investment.