Merton MDP Model

Raphael Abbou

March 10, 2020

1 Model

Continuous MDP model:
States: (t, W_t):
Actions: (π_t, c_t)
Rewards: $\forall t \leq T, U(c_t)$ and $U(T) = U(W_T)$

In this example, we take the utility function $U(x) = -\frac{e^{-ax}}{a}$

2 Discretisation

In order to use or algorithmic MDP resolution, we need to make the problem discrete.

We make the time discrete with timesteps $t = 0, 1, ..., T$. The variance of the Brownian Motion is $T\sigma^2$, Therefore 99.7% of the time, W_T will lie in $[W - 3\sqrt{T}\sigma, W + 3\sqrt{T}\sigma]$.

We consider this interval $[L, U]$ to be the boundaries of W_t, where $L = \max(0, W - 3\sqrt{T}\sigma)$ and $U = W + 3\sqrt{T}\sigma$, and we consider the following discrete space:

$$
\begin{aligned}
&(0, W), \\
&(0, \epsilon), (0, 2\epsilon), ...(0, n\epsilon = U) \\
&(1, \epsilon), 10, 2\epsilon), ...(1, n\epsilon = U) \\
&... \\
&(T, \epsilon), (T, 2\epsilon), ...(T, n\epsilon = U)
\end{aligned}
$$

where ϵ is the discretization step of the price space.

We also make the space of our actions discrete:

$$
\begin{aligned}
\pi_t &= 1\%, 2\%, ...100\% \\
c_t &= 1\%, 2\%, ...100\%
\end{aligned}
$$

Furthermore, we have:

$$
\begin{aligned}
W_{t+1} &= \pi_t(1 + R) + (W_t - \pi_t)(1 + r) \\
&= \pi_t(R - r) + W_t(1 + r)
\end{aligned}
$$

Therefore, we can get the transition probabilities for the MDP. We denote $s_k = L + k\epsilon$ a level of the price discretisation (a state would be (t, s_k)) for all $t < T$

$$\begin{aligned}
\mathbb{P}(s_k \leq W_{t+1} < s_{k+1}) &= \mathbb{P}(s_k \leq \pi_t(R-r) + W_t(1+r) < s_{k+1}) \\
&= \mathbb{P}(\frac{s_k - W_t(1+r)}{\pi_t} + r \leq R < \frac{s_{k+1} - W_t(1+r)}{\pi_t} + r) \\
&= \mathbb{P}(\frac{1}{\sigma}(\frac{s_k - W_t(1+r)}{\pi_t} + r - \mu) \leq \frac{R-\mu}{\sigma} < \frac{1}{\sigma}(\frac{s_{k+1} - W_t(1+r)}{\pi_t} + r - \mu)) \\
&= \Phi(\frac{1}{\sigma}(\frac{s_{k+1} - W_t(1+r)}{\pi_t} + r - \mu)) - \Phi(\frac{1}{\sigma}(\frac{s_k - W_t(1+r)}{\pi_t} + r - \mu)) \\
&= \mathbb{P}^{\pi_t}((t, W_t) \to (t+1, W_{t+1}))
\end{aligned}$$

where Φ is the cdf of the $\mathcal{N}(0,1)$ R.V. At $t = T$, we consider the state to be absorbing.

The rewards are simply defined by:

$$\begin{aligned}
\forall t < T-1, R^{c_t}(t, W_t) &\to (t+1, W_{t+1}) = U(c_t) \\
R^{c_t}(T-1, W_{T-1}) &\to (T, W_T) = U(c_{T-1}) + U(W_T)
\end{aligned}$$