Portfolio Management using Partially Observable Markov Decision Process

Ramin Zahedi and Edwin K. P. Chong

Electrical & Computer Engineering

Colorado State University

Knowledge to Go Places

Definitions

- $p_{i,k}$: Price of stock i on day k
- $n_{i,k}$: Number of shares for stock i on day k
- $r_{i,k}$: Expected return of stock i on day k

$$r_{i,k} = \frac{p_{i,k+1}}{p_{i,k}}$$

- rf_k^{end}: Amount of wealth invested in the riskless asset (bank) at the end of day k
- s : Interest rate of the riskless asset
- rf_k : Amount of wealth invested in the riskless asset (bank) at the beginning of day k

$$rf_k = s\, rf_{k-1}^{end}$$

- W_k: Amount of wealth at the beginning of day k

$$W_k = \sum_{i=1}^{N} n_{i,k} p_{i,k} + rf_k$$

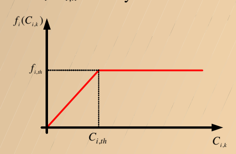
- $\alpha_{i,k}$: Fraction of wealth invested in stock i on day k (the decision)
- $C_{i,k}$: Transaction amount for stock i on day k

$$C_{i,k} = \left|\alpha_{i,k}(W_k - T_k) - n_{i,k} p_{i,k}\right|$$

- T_k : Total transaction costs on day k

$$T_k = \sum_{i=1}^{N} f_i(C_{i,k})$$

where $f_i(C_{i,k})$ usually looks like this:



Dynamics of the problem

- Observe the new prices: $P_k = [p_{1,k}, \dots, p_{N,k}]$
- Choose a candidate action: $a_k = [\alpha_{1,k}, \dots, \alpha_{N,k}]$
- Buy and Sell stocks based on the decision:

$$n_{i,k+1} = Fix\left(\frac{\alpha_{i,k}(W_k - T_k)}{p_{i,k}}\right)$$

- Compute the wealth for the next day:

$$W_{k+1} = (W_k - T_k)\left(\sum_{i=1}^{N}(r_{i,k} - s)\alpha_{i,k} + s\right) \qquad (i)$$

Defining the problem as a Partially Observable Markov Decision Process (POMDP)

State space: $s_k = [N_k, P_k, rf_k, R_k, s]$

where,

- $N_k = [n_{1,k}, \dots, n_{N,k}]$
- $P_k = [p_{1,k}, \dots, p_{N,k}]$
- $R_k = [r_{1,k}, \dots, r_{N,k}]$
- s : Riskless asset's rate of interest

R_k is not known at time k , thus we have a POMDP.

Action space: $a_k = [\alpha_{1,k}, \dots, \alpha_{N,k}]$

Observation space:

$$Z_k = P_k = [p_{1,k}, \dots, p_{N,k}]$$

State transition law: $s_{k+1} = h(s_k, a_k, v_k)$

where,

- h represents the dynamics of the state.

$$h(s_k, a_k, v_k) = [h_1(a_k, N_k, P_k, rf_k), h_2(R_k, v_k)]$$

$$h_1(a_k, N_k, P_k, rf_k) = \begin{bmatrix} N_{k+1} \\ rf_{k+1} \end{bmatrix}$$

- $n_{i,k+1} = Fix\left(\frac{\alpha_{i,k}(W_k - T_k)}{p_{i,k}}\right)$
- $rf_{k+1} = s\, rf_k^{end}$

$h_2(R_k, w_k)$ depends on the model we choose for the price data (ARMA, GARCH, etc.).

- v_k is the uncertainty in the state transitions.

Observation law: $Z_k = [0,1,0,0,0]\, s_k^{tran.}$

Cost function: $g(s_k, a_k) = U(w_{k+1}) - U(w_k)$

where for w_{k+1}, (i) should be used and $U(x)$ is the utility function.

- b_k : The probability distribution over State space S

$$b_k(s) = \Pr.(S_k = s)$$

- $r(b_k, a_k)$: Cost function for the belief-state

$$r(b_k, a_k) = \sum_{s \in S} b_k(s) g(s, a_k)$$

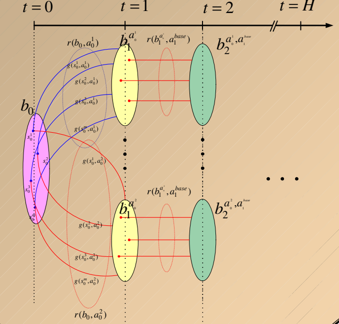
Objective function:

$$V_H = \mathop{E}_{b_1, b_2, \dots}\left[\sum_{k=0}^{H-1} r(b_k, a_k) + U(W_0)\right]$$

The goal:

$$\max_{a_0, a_1, \dots, a_H} V_H$$

Rollout approximation



Curse of dimensionality

- As we move forward in time, the computation increases exponentially (see figure below).
- This problem with POMDP, forces us to approximate V_H, instead of finding the exact value.
- **Q-value:**

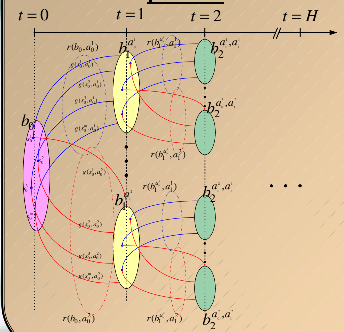
$$Q(b,a) = r(b,a) + \mathop{E}_{b'}\left(V^{\max}(b') \mid b, a\right)$$

- **Approximating Q-value (Rollout):**

Instead of finding the optimal policy π^* for $V^{\max}$, we use a candidate policy called the base policy π_{base} (figure above).

$$Q^{base}(b,a) = r(b,a) + \mathop{E}_{b'}\left(V^{\pi_{base}}(b') \mid b, a\right)$$

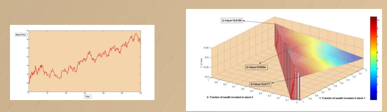
Exact solution for the POMDP problem



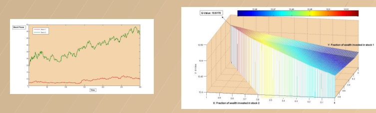
Preliminary results

In all cases mentioned below, the base policy is to invest in the stock that has the maximum rate of return in one period.

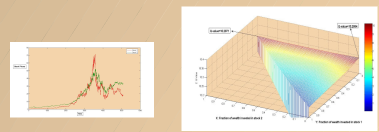
- **Case 1**: Two risky stocks are actually the same. The decision is to divide the wealth equally between the two stocks.



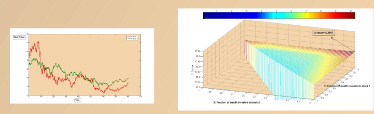
- **Case 2**: Over 300 days, one stock is going up and the other one is not changing. The decision is to invest in rising stock.



- **Case 3**: Over 6000 days, both stocks have a random behavior. As can be seen there's not much difference than putting the money in the bank (investing in none of the stocks) and investing in stock 2.



- **Case 4**: Over 800 days, both stocks are going down. The decision is to put all the money in the bank.



References

- Dimitri P. Bertsekas. "Dynamic Programming and Optimal Control". Athena Scientific, 1995.
- David Blake. "Financial Market Analysis". Wiley, 2000.
- Y. He and E. K. P. Chong, "Sensor scheduling for target tracking: A Monte Carlo sampling approach" *Digital Signal Processing*, vol. 16, no. 5, pp. 533-545, September 2006.
- Harry Markowitz. "Portfolio selection". *The Journal of Finance*, vol. 7, issue 1, pp. 77-91, 1952.
- P Balduzzi, AW Lynch. "Transaction costs and predictability: some utility cost calculations" *Journal of Financial Economics*, vol. 52, pp. 47-78, 1999.