多期間消費投資モデルにおける強化学習を用いたポートフォリオ戦略

横井　稜

2018 年 7 月 9 日

富山県立大学　情報基盤工学講座

July 9, 2018

はじめに

発表の流れ

まえがき

1. ファイナンス工学の主な課題の一つに資金選択問題があり多くの数理モデルは 1 期間を対象としている
2. しかし、現実の市場で多期間資産選択問題についてのモデルを構築するのは難しい
3. そこで、各エージェントが強化学習を利用してポートフォリオ戦略を導出する人工市場を構築する
4. そして、人工市場において構築された多期間消費投資モデルにおけるポートフォリオ戦略がどのような特性を持つか分析

多期間消費投資モデルの概要（１）

市場には無危険資産 $(i = 1)$ と危険資産 $(2 \leq i \leq M)$ が存在し、市場を以下のように仮定

仮定（１）

1. 取引手数料、配当及び税金がない
2. 投資家の投資量は任意の実数
3. 投資家は価格および収益に影響を与えない
4. 投資家は空売りにより収益を達成できる

$$r_{1t} \geq 0 \quad (t = 1, 2, \cdots). \quad \mathrm{E}[r_{it}] \geq \delta + r_{1t} \quad (\delta \geq 0, \exists i, t). \quad \mathrm{E}[r_{it}] \leq K \quad (\forall i, t).$$

r_{1t} は t 期の利子率、r_{it} は t 期に投資機会 $i(i = 2, \cdots, M_t)$ について資本 1 単位から得られる収益（確率変数）を表す。もし期初に第 i 機会に θ を投資した場合、期末には $(1 + \gamma_{it})\theta$ を得る。M_t は t 期に利用可能な投資機会数 $(M_t \leq M)$ である。

また、（非定常）収益率分布 F_t は、各期において独立、あるいはマルコフ連鎖に従うと仮定する。

$$F_t(z_2, z_3, \cdots, z_{M_t}) \equiv Pr\{r_{2t} \leq z_2, r_{3t} \leq z_3, \cdots, r_{M_t t} \leq z_{M_t}\}.$$

多期間消費投資モデルの概要（２）

任意の t 及び空売りが不可能な投資機会 i について $\theta_i \geq 0$ かつ $\sum_{i=2}^{M_t} |\theta_i| = 1$ である任意の θ について収益が負となる可能性の存在条件（下式）が満たされるとする

$$Pr\{\sum_{i=2}^{M_t}(\gamma_{it} - \gamma_{1t})\theta_i < \delta_1\} > \delta_2.$$

意思決定時点 t (t 期末) における資本の投入量 w_t において支払可能制約（下式）が満たされるものとする

$Pr\{w_t \geq 0\} = 1 \quad (t = 1, 2, 3, \cdots, T-1)$

z_{1t} は t 気の貸出金量、z_{it} は t 期初での投資機会 $i(i = 2, \cdots, M_t)$ への投資量、c_t は t 期初に消費に振り向けられる量

$\sum_{i=1}^{M_t} z_{it} = w_{t-1} - c_t$

下式は基本差分方程式で、投入量を求める

$w_t = \sum_{i=2}^{M_t}(r_{it} - r_{1t})z_{it} + (1 + r_{1t})(w_{t-1} - c_t) + y_t \quad (t = 1, 2, 3, \cdots, T)$

下式は、投資家の目的である期待効用を最大化することを表す

$\max \mathrm{E}[U(c_1, \cdots, c_T)]$

多期間消費投資モデルの概要（３）

仮定（２）

1. 個人の寿命（計画期間）は認知である
2. 利子率が決定論的である
3. 労働収入は決定論的である
4. 効用関数が時間加法的である

下式は、仮定の労働収入は決定論的であることを示す。

$$Y_{t-1} \equiv \frac{y_t}{r_{1t}} + \cdots + \frac{y_T}{(1+r_{1t})\cdots(1+r_{1T})}$$

また、効用関数が時間加法的なので以下の式となる。

$$U(c_1, \cdots, c_T) = u_1(c_1) + \alpha u_2(c_2) + \cdots + \alpha^{T-1} u_T(c_T)$$

よって、期間 t における効用関数は以下となり、γ は効用関数の選好の程度を表す。

$$u_t(c_t) = \frac{1}{\gamma} c_t^{\gamma} \qquad (0 < \gamma < 1)$$

人工市場の概要（１）

t 期におけるエージェント n の資産 i への投資量を z_{it}^n、消費に振り向けられる量を c_t^n、また企業からの配当金を d_{it} とすると、投資価値は以下の式となる。

$w_t^n = \sum_{i-2}^{M_t^n}(1+r_{it})z_{it}^n + (1+r_{1t})z_{1t}^n + y_t^n + \sum_{i=2}^{M_t^n}(r_d d_{it}\frac{z_{it}^n}{p_{i(t)}} + z_{it}^{'n})$

r_d は正の定数、$z_{it}^{'n}$ は取引不成立の場合の返金である。収益 r_{it} は下式となる

$r_{it} = \frac{p_{i(t)}}{p_{i(t-1)}} - 1$

配当は、離散の有色ノイズであるとして下式で与える。

$\log\frac{d_{it}}{\overline{d}_{it}} = \epsilon_i^a \log\frac{d_{i(t-1)}}{\overline{d}_{it}} + \epsilon_i^b \xi_{it}$

ξ_{it} は平均 0、分散 σ_i^2 のガウスノイズ、$\epsilon_i^a, \epsilon_i^b$ は $(\epsilon_i^a)^2 + (\epsilon_i^b)^2 = 1$ を満たす正のパラメータ

$\log\frac{d_{it}}{\overline{d}_{it}}$ は平均 0、分散 σ_i^2 で、相関時間 $\tau_s = \frac{1}{\log(\epsilon_i^a)}$ で自己相関関数が減衰する

全資産が一定のもとで組替えを行い、

$\sum_{i=1}^{M_t^n} z_{it}^n = w_{t-1}^n - c_t^n$

市場における危険資産の量はそれぞれ一定であるとする。

$\sum_{n=1}^{N} z_{it}^n = Z_i \quad (2 \le i \le M)$

人工市場の概要（２）

危険資産 i に関する全エージェントが希望する売買量は下式となる。

$B_{it} = \sum_{n=1}^{N} b_{it}^{n}$ $O_{it} = \sum_{n=1}^{N} o_{it}^{n}$

そして $B_{it} \neq O_{it}$ の場合における各エージェントの危険資産の保有量は下式となる。

$z_{it}^{n} = z_{i(t-1)}^{n} + \frac{V_{it}}{B_{it}} b_{it}^{n} - \frac{V_{it}}{O_{it}} o_{it}^{n}$

価格 p_{it} 次のように決定する。

$p_{i(t+1)} = \frac{2p_{it}}{1+\exp\{\frac{-U_{it}}{T_i^P}\}}$　　ただし、$U_{it} = \log \frac{B_{it}}{O_{it}}$

T_i^P は危険資産 i の感度を表す正の定数であり、小さいと需要と供給の差に敏感であり、大きいと鈍感であることを表す。

Bellman 最適方程式の作成（１）

人工市場において、時刻 t にエージェント n が有限の資産のもとでリスクを考慮して各自の判断に従い行動するときの利得を下式で表す。
$V_t^n = \sum_{k=0}^{\infty} (\alpha_n)^k u_{t+k}(c_{t+k}^n)$
α_n はエージェント n のもつ割引率である。

また以下の Bellman 方程式を満たす
$V_\pi^n(s) = \mathrm{E}_\pi[\sum_{k=0}^{\infty} (\alpha_n)^k u_{t+k}(c_{t+k}^n) \mid s_t^n = s]$
$= \sum_a \pi(s,a) \sum_{s'} P_{ss'}^a \{R_{ss'}^a + \alpha_n V_\pi^n(s')\}$
ここで、政策 $\pi(s,a)$ は状態 $s \in S$ において行動 $a \in A(s)$ をとる確率、$P_{ss'}^a$ は政策 $\pi(s,a)$ で状態が s から s' へ遷移する確率、$R_{ss'}^a$ は政策 $\pi(s,a)$ で状態が s から s' へ遷移したときの期待利得である。

また下式の行動価値関数を定義する
$O_\pi^n(s,a) = \mathrm{E}_\pi[\sum_{k=0}^{\infty} (\alpha_n)^k u_t(c_t^n) \mid s_t^n = s, a_t^n = a]$
$= \pi(s,a) \sum_{s'} P_{ss'}^a \{R_{ss'}^a + \alpha_n V_\pi^n(s')\}$

Bellman 最適方程式の作成（２）

上の式は最適政策 π^* に対する状態価値関数、下の式が行動価値関数である。

$$V_*^n(s) = \max_a V_\pi^n(s) \qquad (\forall s \in S)$$

$$Q_*^n(s,a) = \max_\pi Q_\pi^n(s,a) \qquad (\forall s \in S,\ \forall a \in A)$$

よって、最適な価値関数に対する Bellman 方程式である Bellman 最適方程式は以下となる。

$$\begin{aligned} V_*^n(s) &= \max_a Q_*^n(s,a) \\ &= \max_a \sum_{s'} P_{ss'}^a \{R_{ss'}^a + \alpha_n V_\pi^n(s')\} \end{aligned}$$

この Bellman 最適方程式を近似的に解くための手法の 1 つに強化学習がある。

そこで本研究では、強化学習に Actor-Critic モデルを実現するニューラルネットワークを適用し人工市場におけるポートフォリオ戦略を求める。

Actor

エージェント n は環境において状態 x_t^n を観測し、Actor は制御出力を下式で生成

$$q_{it}^n = f\left(\sum_{j=1}^{N_A} W_{ijt}^{A_n} g_j^{A_n}(\mathbf{x}_t^n) + rnd_{it}\right)$$

$$g_j^{A_n}(\mathbf{x}_t^n) = \exp\left\{-\frac{1}{2}(\mathbf{x}_t^n - \mathbf{m}_j^{A_n})^{\mathrm{T}} \mathbf{C}_j^{A_n^{-1}} \times (\mathbf{x}_t^n - \mathbf{m}_j^{A_n})\right\}$$

g_j^A は j 番目の動径基底関数、N_A は Actor の動径基底関数の数、W_{ijt}^{An} は結合荷重、rnd_{jt} は正規乱数である
f はシグモイド関数である。T_n^A はエージェント n の感度である。

$$f(x) = \frac{1}{1 + \exp\{-x/T_n^A\}}$$

Actor の制御出力からエージェント n は下式で取引を行おうとする。

$$b_{it}^n = q'^n_{it} w_t^n - z_{i(t-1)}^n \qquad (q'^n_{it} w_t^n > z_{i(t-1)}^n)$$

$$o_{it}^n = z_{i(t-1)}^n - q'^n_{it} w_t^n \qquad (q'^n_{it} w_t^n < z_{i(t-1)}^n)$$

ここで $q_{it}^{`n}$ は以下で定義

$$q'^n_{it} = \frac{q_{it}^n}{\sum_{i=1}^N q_{it}^n}$$

Critic(1)

Critic は下式で評価値を生成する。N_C は Critic の動径基底関数の数である

$$V_\pi^n(\mathbf{x}_t^n) = \sum_{j=1}^{N_C} W_{jt}^{C_n} g_j^{C_n}(\mathbf{x}_t^n)$$

動径基底関数は Actor で用いたものと同様に、下式で与える

$$g_j^{C_n}(\mathbf{x}_t^n) = \exp\left\{-\frac{1}{2}(\mathbf{x}_t^n - \mathbf{m}_j^{C_n})^{\mathrm{T}} \mathbf{C}_j^{C_n^{-1}} \times (\mathbf{x}_t^n - \mathbf{m}_j^{C_n})\right\}$$

行動の結果、Critic は環境から報酬を受け取り繊維後の状態 x_{t+1}^n を観測する。

$$reward_t^n = u_t(c_t^n) = \frac{1}{\gamma_n}\left(w_{t-1}^n - \sum_{i=1}^{M_t^n} z_{it}^n\right)^{\gamma_n}$$

さらに強化信号として、時刻 t における期待効用と実際の効用との差である TD 誤差 δ_t を Actor へ伝える

$$\mathrm{E}[u_t(c_t^n)] = V_\pi^n(\mathbf{x}_t^n) - \alpha_n V_\pi^n(\mathbf{x}_{t+1}^n)$$

$$\begin{aligned}\delta_t &\equiv u_t(c_t^n) - \mathrm{E}[u_t(c_t^n)] \\ &= u_t(c_t^n) + \alpha_n V_\pi^n(\mathbf{x}_{t+1}^n) - V_\pi^n(\mathbf{x}_t^n)\end{aligned}$$

Critic(2)

また、Critic は同時に、活性度の履歴を計算し

$$e_{jt}^{n} = \lambda_e e_{j(t-1)}^{n} + g_j^{C_n}(\mathbf{x}_t^n)$$

結合荷重を更新する

$$W_{jt}^{C_n} = W_{j(t-1)}^{C_n} + \eta_C \delta_t e_{jt}^{n}$$

Actor では、結合荷重を下式で更新する。η_A, η_C は学習率、λ_e は履歴の減衰率を表す

$$W_{ijt}^{A_n} = W_{ij(t-1)}^{A_n} + \eta_A \delta_t g_j^{A_n}(\mathbf{x}_t^n) \times rnd_{it}$$

シミュレーション方法及び結果（１）

Actor と Critic の動径基底関数はともに各変数を-0.5 から 2.5 の間を 7 等分した値の組合せに中心 m_j^{An}, m_j^{Cn} を配置。C_j^{An}, C_j^{Cn} はすべて 7 行 7 列の単位行列とした。
取引回数を 2100 回とし、最後の 100 回分を 1 試行として計測した。各エージェントの初期の試算は $1000 + 50\sigma$ 書く危険資産の初期の価格は $100 + 5\sigma$ で与えた。また Actor と Critic の初期の結合荷重はともに σ で与えた。配当と労働収入は常に 0 であると仮定した。
まずエージェント数 9 人、価格の感度はすべての危険資産について $T_i^P = 50, (i = 1, 2, 3)$, エージェントの感度はすべて $T_n^A = 1, (n = 1, 2, 3, \cdots, 9)$ とし、その他のパラメータを表 1 のように設定した場合の危険資産の価格と期間 t における効用関数の値の変動の一例が下図である。

表 1　パラメータの値
Table 1　Values of parameters.

η_A	η_C	λ_e	r_{1t}	α	γ
0.001	0.001	0.8	0.01	0.5	0.5

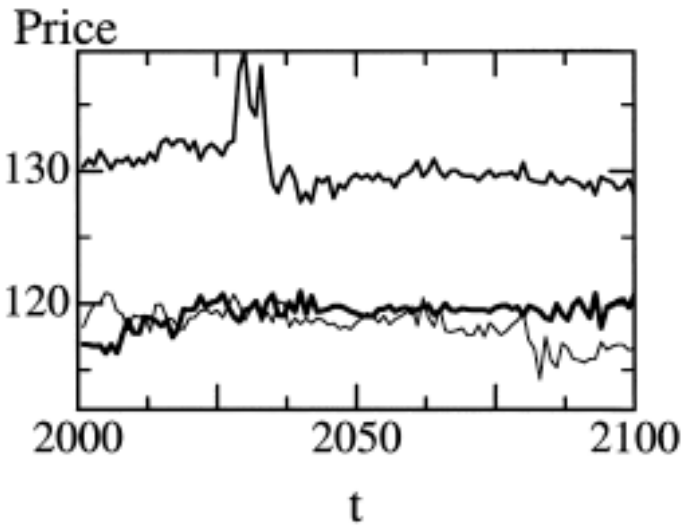

図 3　危険資産の価格の変動

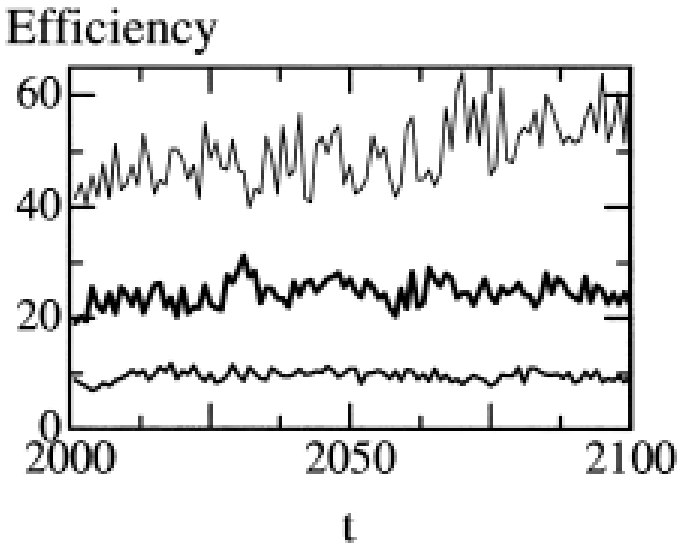

図 4　効用関数の値の変動

シミュレーション方法及び結果（２）

次に、表 1 の将来の報酬に対する割引率と効用関数の選好の程度の部分を変えてその影響について調べる。組合せは $(\alpha, \gamma) = (0.1, 0.1), (0.1, 0.9), (0.5, 0.5), (0.9, 0.1), (0.9, 0.9)$ について、得られた危険資産の価格と効用関数の値が下図である。

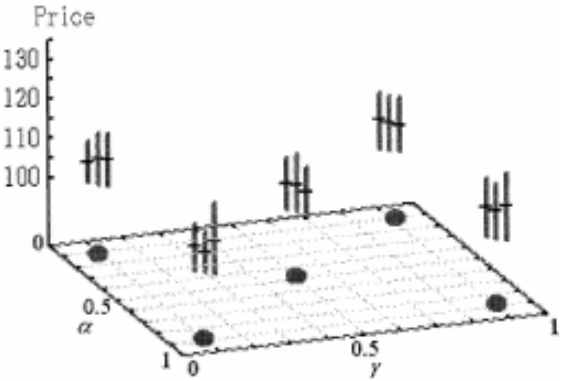

図 5　すべてのエージェントが同じ α や γ の値をもつ場合の価格

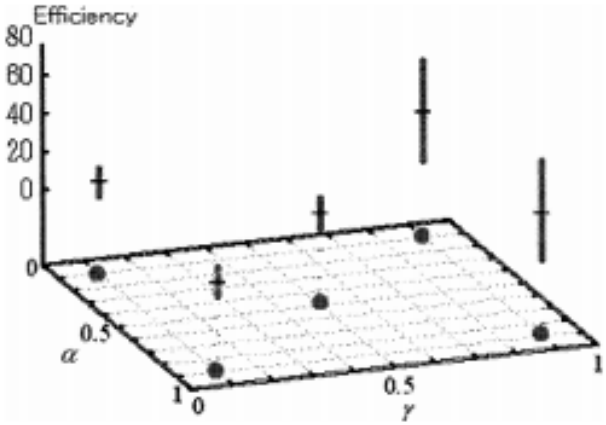

図 6　すべてのエージェントが同じ α や γ の値をもつ場合の効用

この結果から、α が危険資産の価格や効用関数の値に大きな影響を与えないのに対して、γ は大きくなると危険資産の価格と効用関数の値が増加し、併せて効用関数の分散が大きくなることがわかる

シミュレーション方法及び結果（３）

次に、エージェントが異なる割引率と選好の程度の組み合わせを持つ場合についてシミュレーションした。表にはその時のパラメータを表す。

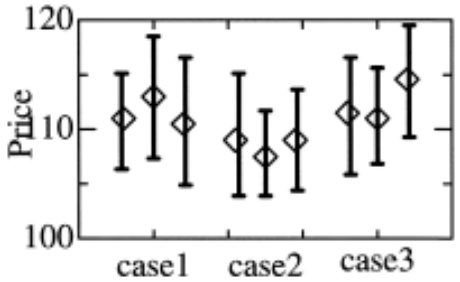

図 7　危険資産の価格

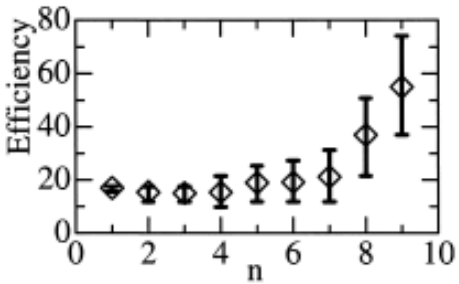

図 8　エージェントの効用（ケース 1）

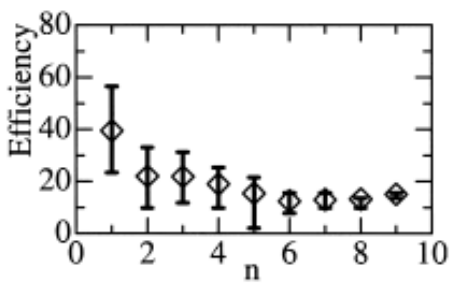

図 9　エージェントの効用（ケース 2）

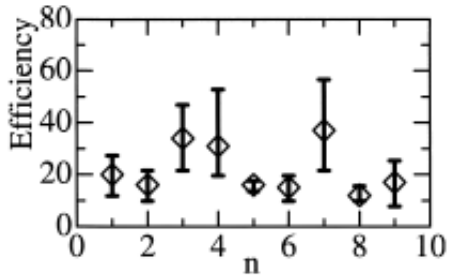

図 10　エージェントの効用（ケース 3）

表 2　α と γ の組合せ
Table 2　Set of α and γ.

	ケース 1		ケース 2		ケース 3	
	α	γ	α	γ	α	γ
1	0.1	0.1	0.1	0.9	0.4	0.6
2	0.2	0.2	0.2	0.8	0.7	0.5
3	0.3	0.3	0.3	0.7	0.2	0.9
4	0.4	0.4	0.4	0.6	0.8	0.7
5	0.5	0.5	0.5	0.5	0.5	0.1
6	0.6	0.6	0.6	0.4	0.1	0.4
7	0.7	0.7	0.7	0.3	0.6	0.8
8	0.8	0.8	0.8	0.2	0.3	0.2
9	0.9	0.9	0.9	0.1	0.9	0.6

これらの結果から、危険資産の価格はそれぞれのケースにおいて大きく異なることがないのにもかかわらず、効用関数の選好の程度を表す γ の値が大きいエージェントほど高い効用関数の値が得られていることがわかる。また、そのばらつきも大きなものとなることも示される。

シミュレーション方法及び結果（4）

次に危険資産に関する価格の感度やエージェントの感度と、取引に参加するエージェントの数を変更した場合に得られた結果を示す。

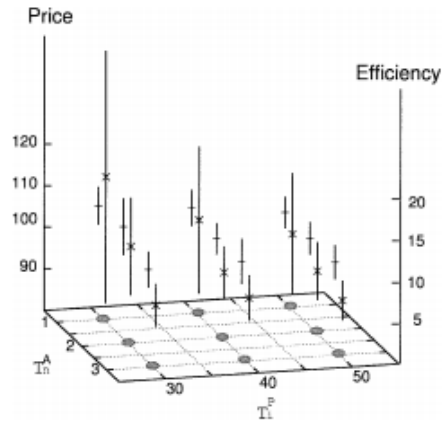

図 11　エージェント数と感度の影響 ($N = 10$)

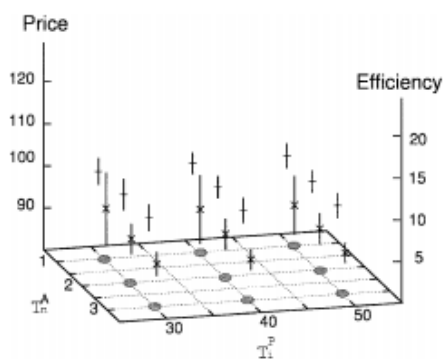

図 12　エージェント数と感度の影響 ($N = 20$)

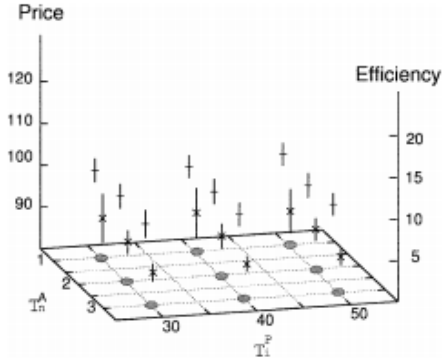

図 13　エージェント数と感度の影響 ($N = 30$)

これらの結果から、危険資産の数に対して市場に参加するエージェントの数が増加すると危険資産の価格や効用関数の値のばらつきが抑制される傾向があることが分かる。
また、危険資産に関する価格の感度が大きくなると、危険資産の価格の値がわずかながら大きくなることがわかる。さらに、エージェントの感度が小さくなるとエージェントの効用関数の値のばらつきが大きくなることがわかる。