

卒業論文（2021 年度）

# ベイジアンネットワークによる株価予測精度の検証

川瀬 一貴 \*

2022 年 1 月

## 概要

本研究では、ベイジアンネットワークによる株価予測について数値実験を通じた予測精度の検証を行った。先行研究である左・北 (2010) や左他 (2011) ではベイジアンネットワークによる株価予測が行われ、いずれも一般的な時系列モデルより精度が高いという結果だった。これらの研究ではベイジアンネットワークの学習に用いたデータと予測の評価に用いたデータの期間が重複していたため、本研究では学習と予測の評価に用いるデータの期間が重複しないように設定した上で、ベイジアンネットワークによる株価予測の精度を改めて検証した。また、さらなる予測精度の向上のため、株価収益率に加えて複数階の対数階差を変数として用いたベイジアンネットワークでの予測も行った。数値実験の結果、今回用いた期間のデータではベイジアンネットワークによる予測が時系列モデルによる予測より高い精度であるとは言えなかった。また、等分割クラスタリングを用いたベイジアンネットワークでは、複数階の対数階差の導入によって予測精度が改善したが、時系列モデルによる予測精度には及ばなかった。ワード法を用いたベイジアンネットワークでは、複数階の対数階差の導入による予測精度の改善はみられなかった。

---

\* 慶應義塾大学経済学部 長倉大輔研究会

# 1 序論

## 1.1 研究背景

従来より株価予測には時系列分析に基づくモデルが代表的なものとして用いられてきた。一方で近年では、機械学習を用いた株価予測の手法が盛んに研究されている。具体例として、時系列データを学習できるニューラルネットワークの一種である LSTM (long short-term memory) を用いた松井他 (2017) や、主に音声認識の分野で使用される DTW (dynamic time wrapping) を改良して用いた中川他 (2017)、さらには主に画像のパターン認識に用いられる CNN (convolutional neural network) を用いた宮崎・松尾 (2017) がある。

これらのような機械学習の手法では、学習のために多くのデータが必要とされる傾向がある。本研究で取り扱うベイジアンネットワークも機械学習手法の一つであり、適切な予測には多くのデータを必要とする。その例として、繁樹他 (2006) では、本研究で用いる Cooper and Herskovits (1992) による手法について、学習に用いるデータ内に全ての変数のパターンが出現していないとき、予測が困難となる場合があることが指摘されている。

しかしながら一般的に、時系列モデルのような統計学的な手法では比較的少ないデータからもある程度の精度で予測することが可能であるとされるが、多くのデータがある場合には機械学習を用いたほうが高い予測精度が出やすいと考えられている。そのため、計算機の性能向上により、従来と比べて多くのデータを使用しても実用的な速度で学習を実行することが可能になった今日においては、先に述べた例のように株価予測の分野でも機械学習の応用が広がっている。

このような時流の中において、ベイジアンネットワークによって株価予測をした先行研究がいくつかある。

左・北 (2010) では、株価収益率を学習に用いたベイジアンネットワークによる予測が、時系列モデルと比較して最大で約 17% 程度、二乗平均平方根誤差 (root mean square error; RMSE) を改善したと報告されている。また、左他 (2011) では、今述べた左・北 (2010) の方法における予測値と実測値との誤差を算出し、その誤差と株価収益率を用いて再度ベイジアンネットワークを構成する手法が実証された。この結果、左・北 (2010) の手法からさらに約 10% 程度、RMSE を改善することに成功した。さらに、原田他 (2012) では、学習に用いる過去の株価収益率の期間を複数設定した上で比較を行い、より過去に遡った株価収益率を学習に含める場合のほうが RMSE が小さくなるという結果が出た。

一方で、これらの先行研究には注意すべき点がある。それは、ここまで紹介してきたベイジアンネットワークによる株価予測を行った先行研究ではいずれも、学習に用いたデータと予測の評価に用いたデータの期間が重複しているという点である。一般的に、学習に用いたデータと同じデータを評価にも用いると、過学習の可能性があり、その手法の汎化性能を正しく評価することができないとされている。

実際に、左 (2012) において株価収益率を変数としたベイジアンネットワークを用いて予測を行う際に、学習に用いたデータと予測の評価に用いたデータの期間が重複している場合は時系列モデルと比較して RMSE が小さくなった一方で、データの期間が重複しないようにした場合は時系列モデルと比較して RMSE が大きくなった。

## 1.2 研究概要

前節にて記した背景を踏まえて、本節では本研究の内容と目的について説明する。

本研究では、前節で言及した左 (2012) による結果を踏まえて、この研究とは別の期間のデータを用い、その上で学習に用いるデータと予測の評価に用いるデータの期間が重複しないようにして数値実験を行う。これによって株価予測におけるベイジアンネットワークの有用性を改めて検証する。

また、株価収益率に加えて 4.2 節で定義する複数階の対数階差を変数として追加したベイジアンネットワークを提案手法とし、これを用いて予測を行う。前節で紹介した原田他 (2012) による結果は、より遡った過去の株価収益率を学習に組み込むほうが良いと示唆していた。このことを受け、株価収益率に加えて、より長期的な株価の傾向を反映するような変数を学習に組み込むことが予測精度の向上に寄与するのではないかという仮説を立てた。したがって、複数階の対数階差を変数として新たに採用し、その効果を検証する。

本研究の目的を改めてまとめると

- (1) 先行研究とは別の期間のデータを用いることで、株価予測におけるベイジアンネットワークの有用性を改めて検証する。
- (2) 複数階の対数階差を変数として追加し、これが予測精度の向上に寄与するかを検証する。

の 2 つである。

## 1.3 本稿の構成

本稿の構成を以下に記す。

まず 2 章では、のちの数値実験において提案手法との比較のために用いる時系列モデルについて紹介する。次に 3 章では、ベイジアンネットワークについて、一般的な概要と予測の際の各手順について順に説明していく。4 章では、3 章で説明した方法を基に、株価収益率の予測を行うための方法として今回使用する提案手法の内容を説明する。そして 5 章では、実際に提案手法を用いて予測を行い、その結果を 2 章で紹介した時系列モデルでの予測の結果と比較する。最後に 6 章で、5 章での数値実験から得られた結果をもとに、本研究のまとめと今後の課題について述べる。

## 2 時系列モデル

本章では、のちの数値実験において提案手法との比較のために用いる時系列モデルとその次数選択の方法について説明する。

### 2.1 AR モデル

本節では、沖本 (2010) に基づいて AR モデル (autoregressive model; 自己回帰モデル) について説明する。

AR( $p$ ) モデルは、値が自身の過去の値によって回帰された形で表現されるモデルであり、 $t$  期の株

価収益率  $r_t$  は,  $p$  期前までの株価収益率  $r_{t-i}$  と, パラメータ  $c, \phi_i$ , 攪乱項  $\varepsilon_t$  を用いて,

$$r_t = c + \sum_{i=1}^p \phi_i r_{t-i} + \varepsilon_t$$

と表される. ただし  $\varepsilon_t$  は, 分散  $\sigma^2$  のホワイトノイズに従う.

## 2.2 MA モデル

本節では, 沖本 (2010) に基づいて MA モデル (moving average model; 移動平均モデル) について説明する.

MA( $q$ ) モデルは, 値が過去のホワイトノイズの線形和で表現されるモデルであり,  $t$  期の株価収益率  $r_t$  は, パラメータ  $\mu, \theta_i$  と,  $q$  期前までの攪乱項  $\varepsilon_t$  を用いて,

$$r_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

と表される. ただし  $\varepsilon_t$  は, 分散  $\sigma^2$  のホワイトノイズに従う.

## 2.3 ランダムウォークモデル

本節では, 沖本 (2010) に基づいてランダムウォークモデル (random walk model) について説明する.

ランダムウォークモデルは,  $t$  期の株価収益率  $r_t$  は,  $t-1$  期の株価収益率  $r_{t-1}$  と, 定数項  $\delta$ , 攪乱項  $\varepsilon_t$  を用いて,

$$r_t = \delta + r_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \sigma^2) \quad (1)$$

と表される. 定数項  $\delta$  はドリフト率と呼ばれ, 式 (1) は, ドリフト率  $\delta$  のランダムウォークと呼ばれることがある.

式 (1) の形から分かる通り, ランダムウォークモデルは, 攪乱項が期待値 0 の iid 系列でありパラメータが  $\phi_1 = 1$  である AR(1) モデルと等しい.

なお, のちの数値実験においては, 参考としてドリフト無しランダムウォークモデルを比較対象に用いることとした.

## 2.4 モデルの次数選択の方法

AR モデルと MA モデルの次数選択において, 赤池情報量規準 (Akaike's information criterion; AIC) を用いた. 本節では, 沖本 (2010) に基づいて AIC について説明する.

AIC は最尤法での推定結果をもとに最良のモデルを選択する客観的な規準であり,

$$\text{AIC} = -2L(\hat{\theta}) + 2k \quad (2)$$

と定義される. ただし,  $L(\hat{\theta})$  はパラメータの最尤推定値における対数尤度,  $k$  は推定したパラメータの数である.

式 (2) の形から分かる通り, AIC は 2 つの項からなる. 第 1 項はモデルの当てはまりを表し, 第 2 項はモデルの複雑さに対するペナルティーを表すと解釈することが可能である. 当てはまりがよくペ

ナルティーが小さいようなモデルが好ましいから、第1項と第2項ともに小さくなるようなモデルが好ましい。したがって、AICを最小にするモデルを選択するのが妥当である。

のちの数値実験ではARモデルとMAモデルについて、モデルの次数として $1, \dots, 5$ のそれぞれを設定して計算を行い、そのうちAICが最小となる値を実際の予測に使用するモデルの次数として選択した。

## 3 ベイジアンネットワーク

### 3.1 ベイジアンネットワークの概要

本節では、繁梲他(2006)、左(2012)を参考にベイジアンネットワークの概要を説明する。

ベイジアンネットワークとは、事象間の関係を確率で表現するグラフィカルモデルである。ベイジアンネットワークは、3つの構成要素、つまり、確率変数、確率変数間の依存関係、依存関係の下での条件付き確率によって定義され、これらをグラフとして表現する。ベイジアンネットワークのグラフでは、確率変数がノードとして表わされる。また、確率変数間の依存関係は各ノード間を結ぶ矢印で表され、この矢印をリンクという。ここで、リンクの元にあるノードを親ノード、リンクの先にあるノードを子ノードという。さらに、各ノード間のリンクで示された依存関係に基づき条件付き確率が計算される。この条件付き確率は、親ノードがある値をとったときに子ノードがある値をとる条件付き確率である。以上のように、ベイジアンネットワークは変数間の確率的な依存関係をモデル化する。

#### 3.1.1 ベイジアンネットワークの簡単な具体例

ここで簡単な具体例を示す。図1と表1が示すベイジアンネットワークは、ある試験で起こりうる事象を構造化して示したものである。まず、3つの事象を以下のように定義する。

- $A$ : 問題集を解く。
- $B$ : カンニングをする。
- $C$ : 合格する。

ここで、 $A, B, C$ をそれぞれ True, False の2つの値をとって事象の生起を表す確率変数とみなす。図1のネットワークは、これら3つの変数をノードとして持ち、それらを結ぶリンクによって依存関係の構造を示している。また、リンクが示す依存関係に基づき、各ノードについての条件付き確率表が表1のように計算され、ノード間の関係が確率によって示される。

以上のように、試験で起こりうる事象についてベイジアンネットワークを用意すると、例えば次のような問題を考えることができる。

学生  $S$  がこの試験を受験して合格した。このとき、学生  $S$  がカンニングをしていた可能性はどの程度だろうか？

この問題は、学生が試験に合格したときにカンニングをしている条件付き確率、つまり  $P(B = \text{True} \mid C = \text{True})$  を求めることによって答えを導ける。この確率は、ベイズの定理を用いて式(3)のように表すことができる。

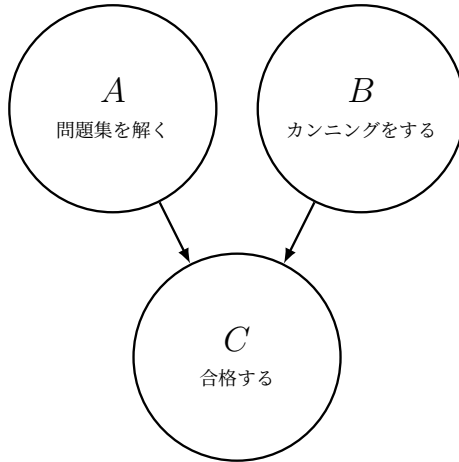


図 1: ある試験に関するベイジアンネットワーク

問題集を解く	
True	False
0.8	0.2

(a) A: 問題集を解く

カンニングをする	
True	False
0.01	0.99

(b) B: カンニングをする

		合格する	
A	B	True	False
T	T	0.9	0.1
T	F	0.7	0.3
F	T	0.6	0.4
F	F	0.1	0.9

(c) C: 合格する

表 1: 各ノードについての条件付き確率表

$$\begin{aligned}
 P(B = \text{True} \mid C = \text{True}) &= \frac{P(B = \text{True}, C = \text{True})}{P(C = \text{True})} \\
 &= \frac{\sum_{A \in (\text{True}, \text{False})} P(B = \text{True}, C = \text{True} \mid A)}{\sum_{A \in (\text{True}, \text{False}), B \in (\text{True}, \text{False})} P(A, B, C = \text{True})} \\
 &= \frac{\sum_{A \in (\text{True}, \text{False})} P(A) P(B = \text{True}) P(C = \text{True} \mid A, B = \text{True})}{\sum_{A \in (\text{True}, \text{False}), B \in (\text{True}, \text{False})} P(A) P(B) P(C = \text{True} \mid A, B)} \quad (3)
 \end{aligned}$$

式 (3) に表 1 に示された条件付き確率を代入して計算すると、問題の答えが求められる。

### 3.2 ベイジアンネットワークによる予測の流れ

本節では、左・北 (2010), 左 (2012) を参考にベイジアンネットワークによる予測の流れについて説明する。ベイジアンネットワークを応用して予測を行う際には、主に

- (1) 確率変数の決定
- (2) ネットワーク構造学習
- (3) 条件付き確率表の獲得
- (4) 確率推論

の4つのステップを踏む。これら各ステップの手法のうち本研究で用いるものについて、具体的な内容を次節以降説明していく。

### 3.3 ネットワーク構造学習

#### 3.3.1 ネットワーク構造の探索アルゴリズム

予測したい対象と予測に用いる変数データを定めたら、ネットワーク構造の学習を行い、グラフのノードとリンクを決定する。左 (2012) によると、ネットワーク構造の探索アルゴリズムとしてよく用いられるものには、全探索アルゴリズム、遺伝的アルゴリズムや K2 アルゴリズムがある。

以下、本小節では 繁樹他 (2006) を参考に、本研究でのネットワーク構造の探索アルゴリズムの選定の経緯を記す。

Chickering et al. (1994) では、ベイジアンネットワーク構造の探索問題は NP 完全問題であることが指摘されている。全探索のための計算量の大きさは、次のような具体例からも分かる。たとえば、変数の数を  $n$  とすると、 $n = 2$  のとき構造数は 3、 $n = 5$  で 29,000、 $n = 10$  でおよそ  $4.2 \times 10^{18}$  となる。そのため、何らかの工夫により探索計算量を減じる必要がある。このことから、本研究では全探索アルゴリズムは採用しない。

また、繁樹他 (2006) によると、遺伝的アルゴリズムは、後述する K2 アルゴリズムの前提として必要とされる変数間の全順序情報がない場合に用いられる手法であり、真の構造を見つけるのが困難な大規模なネットワーク構造の同定では有効であるとされている。しかしながら、本研究で用いる株価収益率のデータは、時間的な全順序関係を持ち合わせている。したがって、本研究では遺伝的アルゴリズムも採用しない。

K2 アルゴリズムは、Cooper and Herskovits (1992) が提唱した、変数間の全順序関係を制約として用いる手法である。繁樹他 (2006) によると、例えば  $X_1 > X_2 > \dots > X_N$  のような変数間の順序関係を前提とすることで、各ノードの親ノードを探す際にその候補数を制限し、より小さい探索空間でネットワーク構造を決定することができる。

本研究で用いる株価収益率のデータは、時間的な全順序関係を持ち合わせており、K2 アルゴリズムを使用するための制約を満たしている。したがって本研究では、先行研究である左・北 (2010)、左 (2012) と同様に、のちの数値実験におけるネットワーク構造学習に K2 アルゴリズムを用いることとした。

#### 3.3.2 K2 アルゴリズム

本小節では、本研究にてネットワーク構造学習に用いる K2 アルゴリズムの具体的な内容について紹介する。図 2 は、Cooper and Herskovits (1992) より参照した K2 アルゴリズムの内容である。

なお、アルゴリズムの 3 行目で用いている  $g(i, \pi_i)$  は、K2 Metric と呼ばれるネットワークの評価指標である。この説明は次小節にて行う。

#### 3.3.3 ネットワーク構造の評価指標

繁樹他 (2006)、左 (2012) によると、一般的に良好なベイジアンネットワークの構造を選択するために用いられる情報量規準は、導出のアプローチから、

- (1) 期待対数尤度からの近似アプローチ

---

**Algorithm 1** K2 algorithm by Cooper and Herskovits (1992)

---

**Input:** A set of  $n$  nodes, an ordering on the nodes, an upper bound  $u$  on the number of parents a node may have, and a database  $D$  containing  $m$  cases.

**Output:** For each node, a printout of the parents of the node.

```
1: for  $i := 1$  to  $n$  do
2:    $\pi_i := \emptyset$ 
3:    $P_{old} := g(i, \pi_i)$  {This function is computed using equation (5).}
4:   OKToProceed := true
5:   while OKToProceed and  $|\pi_i| < u$  do
6:     let  $z$  be the node in  $\text{Pred}(x_i) - \pi_i$  that maximizes  $g(i, \pi_i \cup \{z\})$  {Pred( $x_i$ ) returns the
       set of nodes that precede  $x_i$  in the node ordering}
7:      $P_{new}(x_i) := g(i, \pi_i \cup \{z\})$ 
8:     if  $P_{new} > P_{old}$  then
9:        $P_{old} := P_{new}$ 
10:       $\pi_i := \pi_i \cup \{z\}$ 
11:    else
12:      OKToProceed := false
13:    end if
14:  end while
15:  print ('Node:',  $x_i$ , 'Parents of this node:',  $\pi_i$ )
16: end for
```

---

図 2: K2 アルゴリズム

(2) 予測分布からの近似を行うベイズアプローチ

の 2 つに大別される。

前者の代表例として 2.4 節で説明した AIC がある。AIC は、モデルのデータへの当てはまりの良さに対応する最大対数尤度と、モデルの複雑さに対応するパラメータ数のペナルティーを合わせて評価することで、過学習の回避に寄与している。しかしながら、繁樹他 (2006) によると、AIC は、漸近的一致性の欠如のために真のモデルに対してパラメータ数のより多いモデルを採用してしまう傾向がある。

後者は、モデルの事前分布を所与として、予測分布を求めるアプローチである。BIC のような一貫性を持つ情報量規準は、繁樹他 (2006) によると、一般的にモデルの予測分布を求めたものであり、モデルが特定されれば、そのモデルの予測分布を解析的に求めることができ、より精度の高いモデルを選択することができる。

Heckerman et al. (1995) によると、ディリクレ分布を事前分布とした一般的なベイジアンネット



ワークにおける予測分布を求めると,

$$\text{BD Metric} = \prod_{i=1}^N \prod_{j=1}^M \frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} \prod_{k=1}^L \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})} \quad (4)$$

と表され, これは BD (Bayesian Dirichlet) Metric と呼ばれる.

式 (4) において,  $\Gamma(i)$  は,  $i$  が 1 以上の自然数のとき,  $\Gamma(i) = (i - 1)!$  として計算される. また, 全ノード数を  $N$ , 子ノード  $x_i$  が取りうる状態の総数を  $L$ , 親ノード集合  $Pa(x_i)$  が取りうる状態の総数を  $M$  と表す. さらに, ノード  $x_i$  について, その親ノード集合  $Pa(x_i)$  が状態  $Y^j$  を, ノード  $x_i$  が状態  $X^k$  を取る場合の個数を  $N_{ijk}$  とする.  $N'_{ij}$  と  $N'_{ijk}$  はハイパーパラメータであり,

$$N'_{ij} = \sum_{k=1}^L N'_{ijk}$$

という関係になっている (左 (2012)).

しかし, 一般には事前に知識はなく, 事前分布として無情報事前分布を用いなければならない. ベイズ統計学における無情報事前分布には多くの種類があり, そのために, ベイジアンネットワークでもハイパーパラメータの決定については多くの議論がされていると植野 (2013) は指摘している.

Cooper and Herskovits (1992) は, 一様分布を事前分布としたベイジアンネットワークにおける予測分布を

$$\text{K2 Metric} = \prod_{i=1}^N \prod_{j=1}^M \frac{(L - 1)!}{(N_{ij} + L - 1)!} \prod_{k=1}^L N_{ijk}! \quad (5)$$

と導出した. これは, K2 Metric と呼ばれる.

式 (5) は, 式 (4) にハイパーパラメータ  $N'_{ijk} = 1$  を代入したものに一致することが確かめられ, K2 Metric は BD Metric を特殊化したものであると言える. 繁樹他 (2006) によると, ネットワークのノード数とリンク数が大きいとき, K2 Metric は BD Metric よりも大幅に計算量を減らすことができる.

本研究では, のちの数値実験におけるネットワーク構造学習の際に, ネットワーク構造の評価指標として K2 Metric を用いる.

### 3.4 確率推論

本節では, 繁樹他 (2006), 左 (2012) を基にベイジアンネットワークによる確率推論について説明する.

ベイジアンネットワークの上で行われる各変数の確率分布の計算を確率推論と呼ぶ. ベイジアンネットワークの確率推論では, 観測された確率変数の確定値から, 知りたい確率変数の事後確率を求める. ネットワーク構造が決定され, 変数間の依存関係が明らかになっているとき, 確率変数の確定値  $e$  に対する知りたい確率変数  $x_i$  の事後確率を  $P(x_i|e)$  とすると, 確率推論は以下のような手順で行われる.

- (1) 観測された確率変数の値  $e$  をノードにセットする.
- (2) 知りたい確率変数  $x_i$  について条件付き確率  $P(x_i|e)$  を求める.

知りたい確率変数  $x_i$  の各値についての条件付き確率  $P(x_i|e)$  は、 $x_i$  以外の確率変数についてのありうるすべての状態で平均化する周辺化によって求める。周辺化によれば、 $x_i$  以外の確率変数の確定値  $e$  を得たときに確率変数  $x_i$  が状態  $X_l$  を取る確率  $P(x_i = X_l|e)$  は、

$$P(x_i = X_l|e) = \frac{\sum_{j=1, j \neq i}^N \sum_{x_j=X^1}^{X^L} P(x_1, \dots, x_i = X^l, \dots, x_N, e)}{\sum_{j=1}^N \sum_{x_j=X^1}^{X^L} P(x_1, \dots, x_N, e)}$$

で与えられ、ベイジアンネットワークの条件付き確率表をもとに計算できる。ここで、 $\sum_{x_j=X^1}^{X^L}$  は、確率変数  $x_j$  の取りうるすべての値  $X^1, X^2, \dots, X^L$  について総和を取ることを意味する。

## 4 提案手法

本章では、のちに数値実験にて株価収益率の予測に用いる提案手法の内容について順を追って説明する。

本研究では、左・北 (2010) で行われたベイジアンネットワークによる株価予測の方法を従来手法として、この従来手法に変更を加えたものを提案手法とする。具体的には、株価収益率に加えて 4.2 節で説明する複数階の対数階差を変数として導入することによって、ベイジアンネットワークを構成した。

### 4.1 株価収益率の算出

本研究では、日次の終値の株価データより、株価の対数収益率を計算して予測に用いる。  $t$  期における株価収益率  $r_t$  を

$$r_t = (\ln p_t - \ln p_{t-1}) \times 100$$

と定義する。ここで、 $p_t$  は  $t$  期の株価の終値を表す。

なお、本研究では、予測の対象から 10 期前までの株価収益率を変数として用いた。

### 4.2 複数階の対数階差の算出

前節にて定義した株価収益率に加えて、複数階の対数階差を変数として使用する。  $t$  期における  $n$  期前からの対数階差  $d_{t,n}$  を

$$d_{t,n} = (\ln p_t - \ln p_{t-n}) \times 100$$

と定義する。ここで、 $p_t$  は  $t$  期の株価の終値を表す。

なお、 $t$  期における  $n$  期前からの対数階差  $d_{t,n}$  は、 $t$  期における  $n$  期前からの株価収益率と同等の意味を持つが、前節で定義した  $t$  期における 1 期前からの株価収益率  $r_t$  と明確に区別するため、本研究では複数階の対数階差という名称を用いる。

なお、本研究では予測の対象から 3 期前までの、 $n = 5, 10, 20$  期前からの対数階差を変数として用いた。

### 4.3 変数の離散化

ベイジアンネットワークにおいては、各ノードとなる確率変数が離散値である必要があり、連続値をそのまま予測に用いることはできない。4.1 節、4.2 節で定義した通り、今回予測に用いようとしている変数は連続値をとるため、一度離散化する処理を経由することが必要である。

離散化された株価収益率の取りうる離散値の総数を  $L$ 、離散値を  $r^l$  とする。ここで、離散化された株価収益率  $r^l$  の取りうる値の集合は、

$$\{r^1, r^2, \dots, r^L\} \quad (6)$$

と表される。つまり、連続値である  $t$  期の株価収益率  $r_t$  は、離散化によって集合 (6) の要素に含まれる値に変換されるということになる。

また、後述する方法によって離散化されたクラスタを  $C_l$ 、その重心  $c_l$  をとすると、離散化された株価収益率  $r^l$  が実際にとる値は、各クラスタの重心  $c^l$  であるとする。すなわち、

$$\{r^1, r^2, \dots, r^L\} = \{c_1, c_2, \dots, c_L\}$$

が成立する。

ここまで、株価収益率  $r_t$  の離散化について説明してきたが、複数階の対数階差についても同様に離散化する。

左・北 (2010)、左 (2012) にならって、本研究では、クラスタリングの方法として等分割クラスタリングとウォード法を用いることとした。これら 2 つの方法については、次小節以降で説明する。

また本研究では、離散化された変数取りうる離散値の総数として  $L = 6$  を採用する。これは次のような先行研究の結果に基づいて決定した。離散値総数  $L = 2, 3, \dots, 10$  についてベイジアンネットワークを構成した際の AIC を評価した左 (2012) において、 $L = 6$  で AIC が最小となった。さらに、離散値総数  $L = 2, 3, \dots, 10$  について予測を行った左・北 (2010) において、 $L = 6$  で RMSE (root mean square error) が最小となった。

#### 4.3.1 等分割クラスタリング

等分割クラスタリング (uniform clustering) は、株価収益率のデータを昇順に並べたときの分位数を境界として、各クラスタに含まれる要素数がほぼ均等になるようにクラスタリングする方法である。

#### 4.3.2 ウォード法によるクラスタリング

ウォード法 (Ward's method) は、Ward Jr. (1963) が導入した、クラスタの各要素からそのクラスタの重心までの距離の総和を計算し、それを最小化するようにクラスタリングする方法である。ウォード法では、評価式として、

$$D(C_i, C_j) = E(C_i \cup C_j) - E(C_i) - E(C_j) \quad (7)$$

が用いられる。ただし、

$$E(C_i) = \sum_{z \in C_i} d(z, c_i)^2$$

であり、 $d(z, c_i)$  は  $z$  と  $c_i$  のユークリッド距離を表す。

式 (7) が最小となる 2 つのクラスタの組を 1 つのクラスタとして結合する操作を反復することによって、目標のクラスタ数にクラスタリングすることができる。

#### 4.4 ネットワーク構造学習

本研究では、3.3 節で述べた通り、K2 Metric をネットワーク構造の評価指標として用いた K2 アルゴリズムを採用し、ネットワーク構造を決定する。

#### 4.5 株価収益率の予測

前節までの方法により決定したベイジアンネットワークを  $B$  とし、ネットワーク  $B$  を用いて確率推論を行った結果として得られる、株価収益率が  $r_t = r^l$  となる確率を  $P(r^l|B)$  と定義する。

ここで、 $t$  期の株価収益率の予測値  $\hat{r}_t$  は、

$$\hat{r}_t = \arg \max_{r^l} P(r^l|B)$$

と表される。すなわち、 $P(r^l|B)$  が最大となるような  $r^l$  を株価収益率の予測値として採用するということである。

#### 4.6 まとめ

本章でこれまで説明してきた内容を改めてまとめると、提案手法は、

- (1) 株価収益率  $r_t$  を算出する
- (2) 複数階の対数階差  $d_{t,n}$  を算出する
- (3) 変数を離散化する
- (4) 離散化された変数を用いてベイジアンネットワーク  $B$  の構造学習を行う
- (5) ネットワーク  $B$  によって株価収益率の予測値  $\hat{r}_t$  を算出する

の 5 ステップからなる。

## 5 数値実験

### 5.1 使用するデータ

今回は日経平均株価の日次の終値のデータを使用する。1996 年 10 月 14 日から 2021 年 3 月 31 日までの 6000 日分を用いてベイジアンネットワークを構成し、このネットワークを用いて 2021 年 4 月 1 日から 2021 年 4 月 30 日までの株価収益率を予測した。

### 5.2 用いるモデル

予測精度の比較のため、

- AR モデル

- MA モデル
- ランダムウォークモデル
- 従来手法
- 提案手法

の5つのモデルを用いて予測をする。

ここで、従来手法とは、変数として株価収益率のみを用い、その他の部分は提案手法と同様に予測を行う手法とする。

なお、従来手法と提案手法では、変数の離散化のために等分割クラスタリングを用いる場合とウォード法を用いる場合の両方で予測を行う。

### 5.3 予測の評価方法

それぞれのモデルで行った予測の評価には、二乗平均平方根誤差 (root mean square error; RMSE), 最大誤差 (maximum error; max. error), 最小誤差 (minimum error; min. error), 相関係数 (correlation coefficient; CC) の4つを用いて、予測精度を比較する。それぞれの定義は以下の通りである。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (r_t - \hat{r}_t)^2}$$

$$\text{Max. Error} = \arg \max_t |r_t - \hat{r}_t|$$

$$\text{Min. Error} = \arg \min_t |r_t - \hat{r}_t|$$

$$\text{CC} = \frac{\sum_{t=1}^n (r_t - \bar{r})(\hat{r}_t - \bar{\hat{r}})}{\sqrt{\sum_{t=1}^n (r_t - \bar{r})^2} \sqrt{\sum_{t=1}^n (\hat{r}_t - \bar{\hat{r}})^2}}$$

ただし、 $n$  はデータの組の総数を、 $\bar{r}_t, \bar{\hat{r}}_t$  はそれぞれ  $r_t, \hat{r}_t$  の平均値を表す。

### 5.4 結果

まず、変数を離散化する際にできた各クラスタについて、等分割クラスタリングによる結果は表2、ウォード法による結果は表3に情報をまとめた。なお、各表において、 $c_l$  はクラスタ  $C_l$  に含まれるデータの重心を表す。

クラスタ	区間	データ数	$c_l$
$C_1$	[-12.11%, -1.18%)	1000	-2.24%
$C_2$	[-1.18%, -0.42%)	1000	-0.76%
$C_3$	[-0.42%, 0.04%]	1000	-0.18%
$C_4$	(0.04%, 0.51%]	1000	0.26%
$C_5$	(0.51%, 1.21%]	1000	0.83%
$C_6$	(13.23%, 13.23%]	1000	2.12%

表2: 等分割クラスタリングによる株価収益率のクラスタ

クラスタ	区間	データ数	$c_l$
$C_1$	[-12.11%, -4.35%)	50	-6.11%
$C_2$	[-4.35%, -1.09%)	1041	-1.95%
$C_3$	[-1.09%, -0.07%]	1670	-0.50%
$C_4$	(-0.07%, 1.68%]	2636	0.63%
$C_5$	(1.68%, 3.92%]	555	2.32%
$C_6$	(3.92%, 13.23%]	48	5.49%

表 3: ウォード法による株価収益率のクラスタ

次に、データより構成したベイジアンネットワークのネットワーク構造を、等分割クラスタリングを用いたものは図 3 に、ウォード法を用いたものは図 4 に示した。なお、図 3 と図 4 においてノード内に記されている文字について、例えば  $r_t$  は  $r_t$  を示し、 $d_{t-1.5}$  は  $d_{t-1.5}$  を示す。

最後に、得られたベイジアンネットワークによって株価収益率を予測し、実測値と比較した評価を表 4 と図 5 に示した。参考として他のモデルによる結果と合わせて表示している。ただし、表 4 と図 5 において、RW はドリフト無しランダムウォークモデルを示す。また、Uniform1, Ward1 はそれぞれ等分割クラスタリングを用いた従来手法、ウォード法を用いた従来手法を示し、Uniform2, Ward2 はそれぞれ等分割クラスタリングを用いた提案手法、ウォード法を用いた提案手法を示す。

以下、表 4 と図 5 から読み取れる各モデルの評価について記す。

RMSE が最小となるのは AR(1) モデルであり、ほぼ同等の値で MA(1) モデルが続く。等分割クラスタリングを用いた提案手法は従来手法より RMSE が約 30% 程度小さくなったが、AR(1) モデル・MA(1) モデルには及ばなかった。また、ウォード法を用いた提案手法は従来手法と RMSE が変わらず、これも AR(1) モデル・MA(1) モデルには及ばなかった。

最小誤差については各モデルで大きな差はなかった。一方で、最大誤差についてはベイジアンネットワークを用いた手法が AR(1) モデル・MA(1) モデルと比べて、全体的に約 12% ~ 19% 程度大きくなった。また、等分割クラスタリングを用いた提案手法は従来手法より最大誤差が約 6% 程度小さくなった。

モデル	RMSE	Max. Error	Min. Error	CC
AR(1)	1.016	2.272	0.006	0.154
MA(1)	1.015	2.266	0.006	0.163
RW	1.549	4.411	0.064	-0.154
Uniform1	1.456	2.698	0.045	0.133
Uniform2	1.026	2.541	0.111	NaN
Ward1	1.230	2.684	0.088	0.000
Ward2	1.230	2.684	0.088	0.000

表 4: 各モデルについての予測の評価

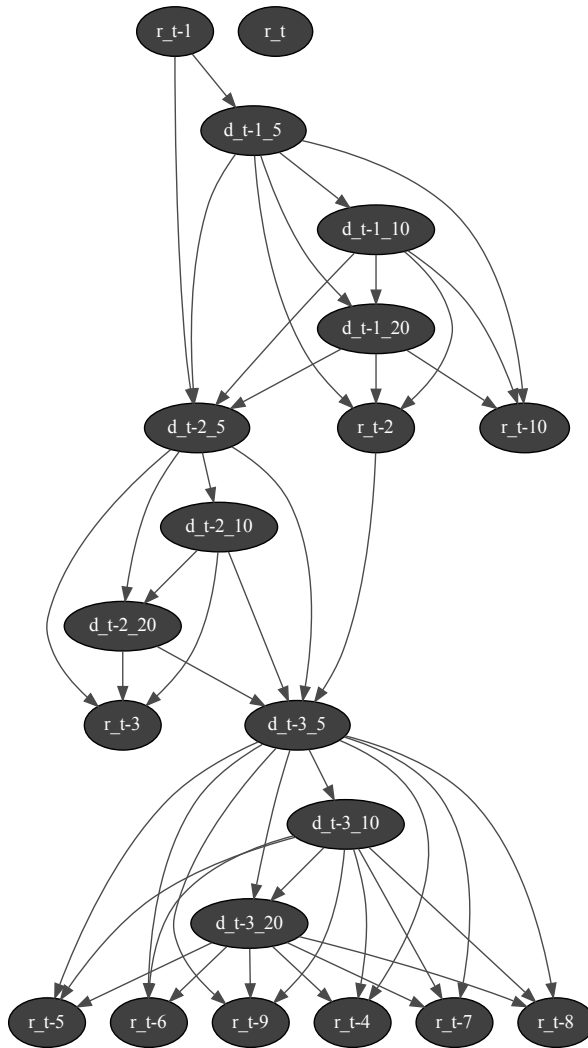


図 3: 等分割クラスタリングを用いたベイジアンネットワーク

## 6 結論

本章では、数値実験から得られた結果をもとに、本研究のまとめと今後の課題について述べる。

### 6.1 まとめ

まず、1.2 節で説明した本研究の目的は、

- (1) 先行研究とは別の期間のデータを用いることで、株価予測におけるベイジアンネットワークの有用性を改めて検証する。

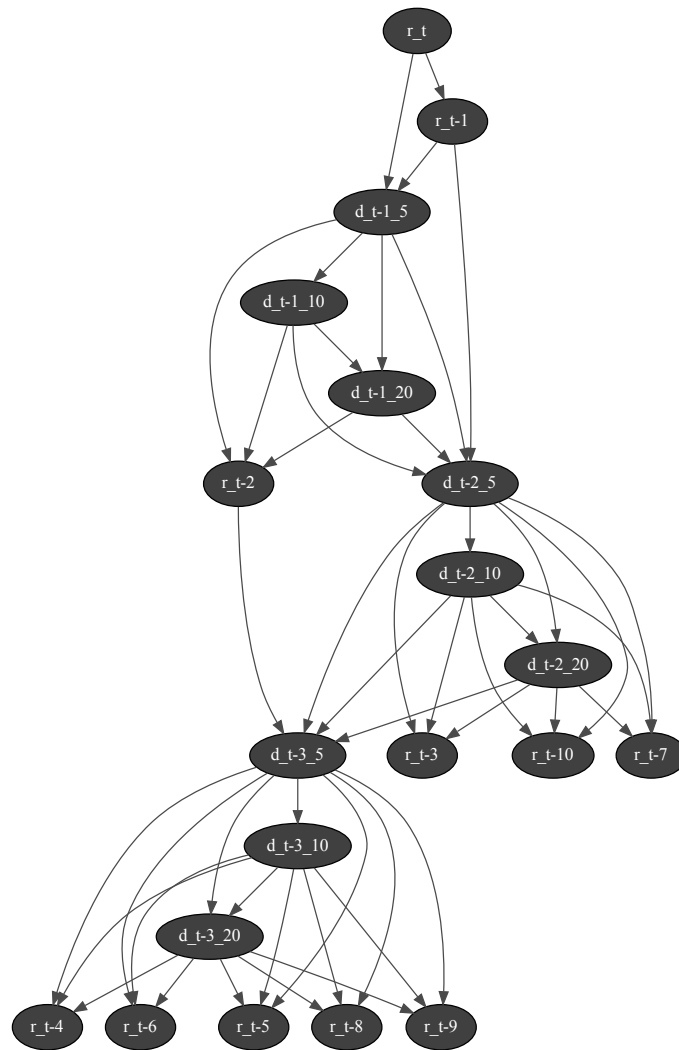


図 4: ウォード法を用いたベイジアンネットワーク

(2) 複数階の対数階差を変数として追加し、これが予測精度の向上に寄与するかを検証する。

の 2 つであった。

1 点目について、本研究で用いた期間についても先行研究 (左 (2012)) と同様に、ベイジアンネットワークによる予測の RMSE が時系列モデルによるものより大きかった。したがって、ベイジアンネットワークによる株価予測が時系列モデルより高精度であるとは言えない。

2 点目について、株価収益率に加えて複数階の対数階差を導入することで等分割クラスタリングを用いたベイジアンネットワークでは RMSE の減少がみられ、予測精度が改善した。しかしながら、時系列モデルでの精度には及ばなかった。一方で、ウォード法を用いたベイジアンネットワークでは複数階の対数階差の導入による予測精度の改善はみられなかった。



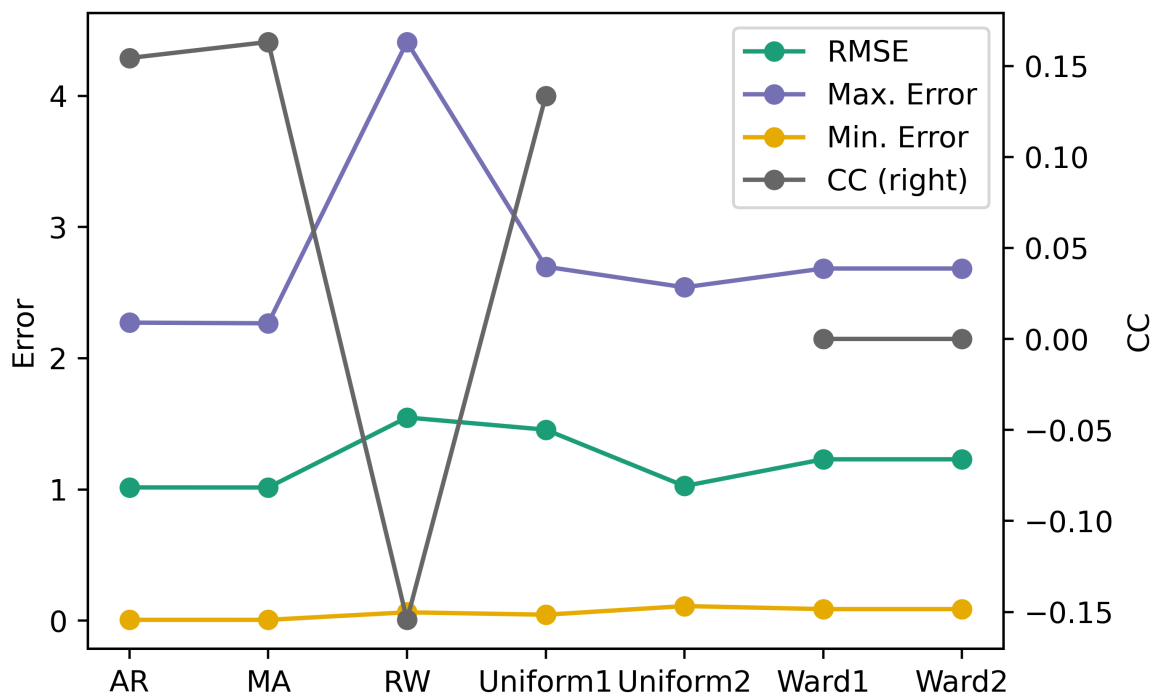


図 5: 各モデルについての予測の評価

## 6.2 今後の課題

図 3 と図 4 から見て取れる通り、今回得られたベイジアンネットワークはいずれも、予測の対象である株価収益率  $r_t$  のノードとリンクで結ばれているノードが少ない。このことから、今回使用したデータについて、ネットワーク構造学習に用いた変数に  $r_t$  と依存関係の大きいものがあまり含まれていなかったことが分かる。したがって、 $r_t$  のノードを起点とするリンクがより多くなるように、 $r_t$  と依存関係の大きい変数を複数学習に組み込めれば、予測精度の改善が見込める可能性がある。このような変数を発見することが今後の課題である。

## 謝辞

本稿の執筆にあたり、慶應義塾大学経済学部の高倉大輔教授には、ゼミ内での発表の際ご指導・ご助言を頂きました。ここに感謝を申し上げます。

## 参考文献

- Chickering, David M., Dan Geiger, David Heckerman et al. (1994) “Learning Bayesian networks is NP-hard,” Technical report, Citeseer.
- Cooper, Gregory F. and Edward Herskovits (1992) “A Bayesian method for the induction of probabilistic networks from data,” *Machine learning*, Vol. 9, No. 4, pp. 309–347.
- Heckerman, David, Dan Geiger, and David M. Chickering (1995) “Learning Bayesian networks: The combination of knowledge and statistical data,” *Machine learning*, Vol. 20, No. 3, pp. 197–243.
- Ward Jr., Joe H. (1963) “Hierarchical grouping to optimize an objective function,” *Journal of the American statistical association*, Vol. 58, No. 301, pp. 236–244.
- 植野真臣 (2013) 『ベイジアンネットワーク』, コロナ社, 東京.
- 沖本竜義 (2010) 『経済・ファイナンスデータの計量時系列分析』, 統計ライブラリー, 朝倉書店.
- 左毅 (2012) 「ベイジアンネットワークの株価予測への応用」, 博士論文, 名古屋大学.
- 左毅・北栄輔 (2010) 「ベイジアンネットワークを用いた株価予測について」, 『情報処理学会論文誌数理モデル化と応用 (TOM)』, 第 3 巻, 第 3 号, 80–90 頁, 10 月.
- 左毅・原田昌朗・北栄輔 (2011) 「ベイジアンネットワークを用いた株価予測法の精度改善」, 『情報処理学会論文誌数理モデル化と応用 (TOM)』, 第 4 巻, 第 4 号, 92–103 頁, 11 月.
- 繁榎算男・植野真臣・本村陽一 (2006) 『ベイジアンネットワーク概説』, 培風館, 東京.
- 中川慧・今村光良・吉田健一 (2017) 「株価変動パターンの類似性を用いた株価予測」, 『人工知能学会全国大会論文集第 31 回全国大会 (2017)』, 2D11–2D11 頁, 一般社団法人人工知能学会.
- 原田昌朗・左毅・北栄輔 (2012) 「ベイジアンネットワークを用いた株式指数予測の改良」, Technical Report 21, 名古屋大学大学院情報科学研究科, 名古屋大学大学院情報科学研究科, 名古屋大学大学院情報科学研究科.
- 松井・藤五郎・汐月・智哉 (2017) 「LSTM を用いた株価変動予測」, 『人工知能学会全国大会論文集第 31 回全国大会 (2017)』, 2D3OS19a2–2D3OS19a2 頁, 一般社団法人人工知能学会.
- 宮崎邦洋・松尾豊 (2017) 「Deep Learning を用いた株価予測の分析」, 『人工知能学会全国大会論文集第 31 回全国大会 (2017)』, 2D3OS19a3–2D3OS19a3 頁, 一般社団法人人工知能学会.