ギブスサンプリング法を用いた構造変化に関する分析：パンデミック前後における投資家の売買取引への適用

村原英樹
北九州市立大学 経済学部 経営情報学科
hmurahara@kitakyu-u.ac.jp

森脇敏雄
北九州市立大学 経済学部 経営情報学科
moriwaki1040@kitakyu-u.ac.jp

概要

我々は，株式市場における投資家の売買取引の回数に着目し，それらがポアソン分布に従うことを仮定したモデルについて考察した．特に，日ごとの売買取引の回数がパンデミック前後で変化したか否かと，変化した場合の構造変化点について検討した．航空業界に属するANA Holdingsを対象に分析を行った結果，次のことが明らかになった．パンデミックの前後を含む期間において，日ごとの売買取引の回数に関して，構造変化が起こっている．その構造変化が起きたのは，緊急事態宣言が出された 2020 年 1 月 30 日よりも後の時点の，2020 年 2 月 21 日である．2020 年 2 月 21 日の直後の期間は，市場全体では株価が低迷している時期であり，航空業界では中国便の運休，国内線の減便といった事象が発生している．本稿の分析結果は，ANA Holdings において，新型コロナウイルスの感染拡大に起因し，日ごとの売買取引の回数に構造変化が起こったことを示唆している．

1　はじめに

本稿の目的は，株式市場における投資家の売買取引について，ベイズ統計学以外の伝統的な手法を用いてデータ分析を行う方に向けて，マルコフ連鎖モンテカルロ法（特に，ランダム化ギブスサンプリング法；以下，ギブスサンプリング法）を用いた新たな分析手法を紹介することである．この分野でギブスサンプリング法を用いた分析をすることは，それほど一般的ではないと思われ，その意味である程度新鮮なものになればと思っている [*1]．可能な限り単純な方法を用い，本稿を読んだだけで，類似の状況下においては，同様の分析がすぐにできるようになることを目標としている．

さて，本稿についての概略を述べよう．本稿では，株式市場における投資家の売買取引について，ギブスサンプリング法を用いた構造変化に関する分析を行う．対象企業は，全日本空輸株式会社（以下，ANA Holdings）とし，対象期間は，2018 年 1 月 1 日から 2023 年 6 月 23 日までとする．なお，対象企業としてANA Holdingsを選択したのは，当該企業の投資家が，新型コロナウイルス感染拡大に起因する環境の変化に直面している可能性が高く，構造変化の分析に適していると考えたためである．対象期間として上記を選択したのは，新型コロナウイルス感染拡大の影響を分析する際に，前後の期間を十分に含むようにするという判断によるものである．

[*1] とはいうものの，少し横の分野（計量経済学などの分野）では，使われることもある手法ではある．例えば [7] の 4 章では類似の手法が紹介されており，本稿の内容に興味を持ってくださった読者は，そちらも併用して読まれることを勧める．

本論文の構成としては，まず第 2 節で本稿で行う考察内容と手順の概略について述べる．次いで第 3 節でポアソン分布について復習する．マルコフ連鎖モンテカルロ法を使用するにあたっては，必要となる統計モデルを仮定する必要があるが，本稿の分析ではポアソン分布を統計モデルとして仮定するため，その妥当性についての議論もここで行う．続いて第 4 節で基礎概念の復習を行う．ギブスサンプリング法の概略をはじめ，確率密度関数など基本的な内容を念のために確認しておく．第 5 節では，ギブスサンプリング法でサンプリングを行うために必要な事項を説明する．本稿ではk, λ, μの 3 種類の値をサンプリングするので，それに関連する方法についてのみ紹介する．特に，ポアソン分布を統計モデルとして仮定する場合，対応する事前分布と事後分布がガンマ分布となることが重要であるが，その辺りの内容について説明する．第 6 節では，第 5 節で説明したギブスサンプリング法によって得られた分布が詳細釣り合い条件を満たし，目標分布に収束することについて述べる．さらに実際のプログラミングにおいて，実行結果が収束しているかの確認を行うために必要な Geweke による収束診断を紹介する．第 7 節では，ANA Holdingsを対象とし，株式市場における投資家の売買取引について，構造変化に関する分析を行う．第 8 節では，目標分布を直接得る方法を紹介する．

2　本稿で行う考察内容と手順の概略

本稿では前節で述べたように，2018 年 1 月 1 日から 2023 年 6 月 23 日までの株式市場における投資家の売買取引について，マルコフ連鎖モンテカルロ法の特殊な場合であるギブスサンプリング法（4.1 節参照）を用いた分析を行う．その際，特に次の 2 つについて考察する．

1. 日ごとの売買取引の回数は，パンデミックの前後で変化したか？
2. 変化したとすれば，その変化点はいつか？

日ごとの売買取引の回数はポアソン分布に従うことを仮定し（3 節参照），全期間をT，パンデミックによって構造変化が起こった日をkで表す．またこの日を境とした前半と後半の日ごとの売買取引の回数の平均をそれぞれλ, μで表す．観測値k, λ, μに対応する確率変数をK, Λ, Mで表し，それらの従う初期の事前分布を以下で仮定する [*2]．

$$K \sim \mathrm{U}(1, T), \quad \Lambda \sim \mathrm{Ga}(\alpha_0, \beta_0), \quad M \sim \mathrm{Ga}(\gamma_0, \delta_0).$$

このとき，このモデルに対する尤度関数は，$x_{1,T} = x_1, \ldots, x_T$として，

$$L(k, \lambda, \mu \mid x_{1,T}) \propto \lambda^{\sum_{t=1}^{k} x_t} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t} e^{-(T-k)\mu}$$

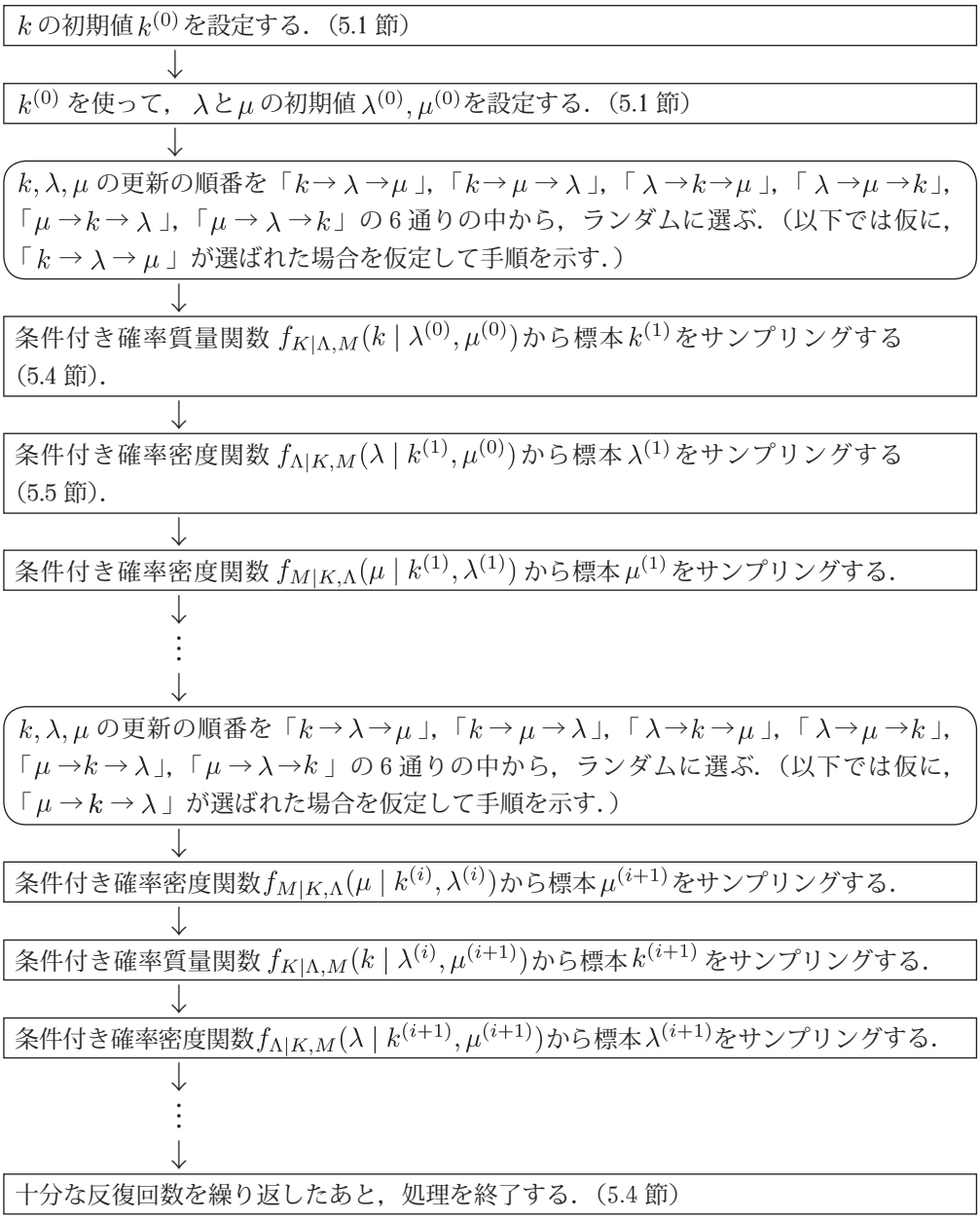
となることが示され，K, Λ, Mの従う同時事後分布は

$$f_{K,\Lambda,M|x_{1,T}}(k, \lambda, \mu \mid x_{1,T}) \propto \lambda^{\alpha_0 - 1 + \sum_{t=1}^{k} x_t} e^{-(k+\beta_0)\lambda} \cdot \mu^{\gamma_0 - 1 + \sum_{t=k+1}^{T} x_t} e^{-(T-k+\delta_0)\mu}$$

となることが示される（5 節参照）．この同時事後分布は，その時々によって，1 つ以外の変数

[*2] 通常ベイズ統計学では，紛れを生じさせないため大文字を使わないが，本稿では 5.1 節に書く方針に依ることにする．

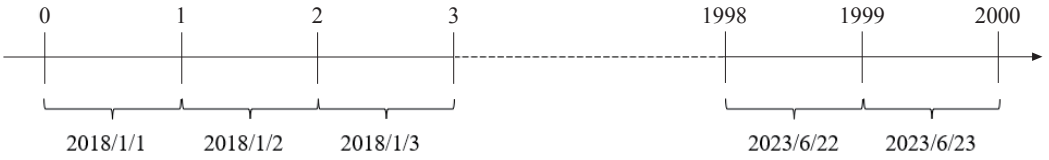
を固定した状態で得られる条件付き確率分布として使用され，ギブスサンプリング法（4.1 節参照）を用いる各ステップで更新され続ける．具体的な手順は下の図の通りであり，詳細は後の節を参照されたい．なお，モデルの妥当性に関する議論を本来は入念に行う必要があるが，本稿では，わかり易さと応用のし易さを優先することにする．



3 ポアソン分布を用いた分析について

本節では，日ごとの売買取引の回数がポアソン分布に従うと仮定することの妥当性について議論する．図 1 のような，長さL の期間を考え，この期間中にランダムにn 個のイベントを配置する状況を考えよう [*3]．（下の図は，$L=2000$ とした図を描いている．）

図 1　考察する対象の時間軸



各単位期間（例えば，$[0,1]$ や $[1,2]$ のような長さが 1 の区間）に 1 つの点が配置される確率は $1/L$ であるから，各単位期間 $[t,t+1]$ に k 個の点が配置される確率は，

$$P(X_t=k)={}_nC_k\left(\frac{1}{L}\right)^k\left(1-\frac{1}{L}\right)^{n-k}$$

となる．今，「期間 L を十分大きくとり，ランダムに配置されるイベントの個数 [*4] n との比 n/L が一定であるような状況」を考えると，その状況下で「単位期間あたりに，k 個のイベントが行われる回数は，母数 n/L のポアソン分布に従う」ことを示そう．説明の簡略化のため，$p=1/L$ とし，$\lambda=n/L=np$ とする．λを一定とすると，上の式は

$$\begin{aligned}
P(X_t=k)&=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\
&=\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}p^k(1-p)^{n-k}\\
&=\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}\\
&=\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-k}
\end{aligned}$$

となる．ここで，

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=\lim_{n\to\infty}\left(\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right)^{-\lambda}=e^{-\lambda}$$

[*3] この期間は考察する対象の時間軸をイメージし，左端の点が考察の開始時刻，右端の点が終了時刻と捉えるとよい．例えば，後で紹介する分析の対象期間は，2018 年 1 月 1 日から 2023 年 6 月 23 日の期間であるため，左端の点は 0 で右端の点はその期間の日数（計算すると 2000 日）となる．また n 個の点をランダムに配置する状況は，この期間にランダムに n 回の取引が行われる状況と同じであると捉えられる．

[*4] イベントの回数とした方が日本語としては自然であると思うが，ここでは個数という呼称を用いることにする．

だから，

$$\lim_{n\to\infty} P(X_t = k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

となる．右辺はポアソン分布の確率質量関数であり，このことから，n/L が一定かつ n（または L）が十分大きいとき，上の状況はポアソン分布で近似できることがわかる．すなわち，単位期間にランダムに起こるイベントの個数は，平均 λ のポアソン分布に（近似的）に従う[*5]．

ここで，上の状況に，株式市場における投資家の売買取引を当てはめてみよう．本稿では考察に先立って「売買取引は，与えられた期間[*6]中に等確率で起こるイベントとして捉える」という仮定をおき[*7]，さらに議論の簡略化のため，以後，L は整数値をとると思うことにする．東京証券取引所においては，平日の 9 時から 11 時 30 分を前場，12 時 30 分から 15 時を後場とし，これらの取引時間内に売買取引が行われている[*8]．したがって，土日・祝日に加えて，平日の取引時間外については，売買取引が行われていない．そのため本稿では，図 1 の時間軸から，売買取引が行われていない日および時間帯を削除し，さらに削除した区間の両端をくっつけて作った新たな期間を考察期間とする[*9]．この新たな期間の長さを，先ほどと同様に L とおき n 回の売買取引が行われたと仮定すると，$np = n/L$ が一定，すなわち単位時間あたりの売買取引の回数が一定，かつ n が十分大きいとき，日ごとの売買取引の回数の分布はポアソン分布で近似できることがわかる．以降の節では，日ごとの売買取引の回数 X_t（ただし，t は日付に対応する値）がポアソン分布に従うことを仮定し，マルコフ連鎖モンテカルロ法を用いた分析を行う．

4 基礎概念の復習

4.1 ギブスサンプリング法について

経済学などの分野，特にベイズ的な計量経済分析などの分野においては，複雑な確率分布からの標本抽出にギブスサンプリング法が用いられることがある．ギブスサンプリング法は，マルコフ連鎖モンテカルロ法の一種で，直接サンプリングすることが難しい確率分布から，目的の確率分布に近い標本列（サンプル）を生成するものである．すなわちこれは，多変量確率分布から一連の標本列を得る方法であって，観測されたデータが与えられたときに，そのデータが従う分布（統計モデル[*10]）と調べたい変数が従う分布（事前分布）を仮定し，それらから（ベイズの定理によって）導かれる事後分布を使って，標本を得る行為を繰り返し行い，それによって得られた標本列をもとに，標本列が従う分布を特定しようとすること（マルコフ連鎖モンテカルロ法）のある特別な場合のことをギブスサンプリング法という．

---

*5 L を整数としても，それが十分長い期間である場合は，その仮定はポアソン分布で近似できることに大きな影響を及ぼさない．

*6 与えられた期間とは，必ずしも全期間とは限らない．

*7 この仮定が妥当であるかどうかについては，後述の分析内容を前提とした上で，読者に判断を委ねたい．

*8 詳細については，日本取引所グループ（JPX）のウェブページ（`https://www.jpx.co.jp/equities/trading/domestic/index.html` を参照してほしい．

*9 本稿で使用するデータでは，売買取引が行われた日数は 1,334 日になる．

*10 統計モデルの正確な定義は，5.2 節で行う．

本稿で考える状況に適したざっくりとした説明をしよう．本稿ではK, Λ, Mが，それぞれどのような分布に従っているのか知りたいのだが，それがわからない．そこでK, Λ, Mが従っていそうな分布（事前分布）を何かしら仮定する．さらに，我々の知りたい分布に近づけるのに妥当だと思われる統計モデル（データが従う分布）を仮定し，統計モデル・事前分布とベイズの定理を使って事後分布を求め（ベイズ更新），さらに事後分布から標本を 1 つサンプリングする．さらに，ここで得られた事後分布を新しい事前分布と思って，先ほどの統計モデルとベイズの定理を使って新しい事後分布・そこからの標本を得る行為を行う．この試みを繰り返す中で，1 回のベイズ更新を繰り返すごとにK, Λ, Mの実現値，k, λ, μ が 1 つずつサンプリングされるのであるが，これをできるだけたくさん集める．そして，その結果（サンプリングされた標本列）をもとに，k, λ, μがどのような分布に従いそうかを考察するのである [*11]．

一般的なマルコフ連鎖モンテカルロ法やギブスサンプリング法についての入門的な解説については [9]，またより詳細な説明は [2]，証明を含んだ内容などに関しては [6] や [13] などを参照されたい．加えて，北九州市立大学経済学部が発行している商経論集にも，関連論文 [5] がある．また本稿で扱っている内容と類似した内容が [7] にあり，この説明はわかりやすいと思われる．

4.2 確率質量関数および確率密度関数などについて

Ωを全事象，$\mathcal{F}$をΩ上の完全加法族，Pを（Ω,$\mathcal{F}$）上の確率（すなわち，Pは$\mathcal{F}$から [0, 1] への写像で所望の条件を満たすもの [*12]）とするとき，$\omega \in \Omega$ に対して実現値$X(\omega) \in \mathbb{R}$を対応させる写像Xを確率変数とよぶ．一般に，任意の実数xに対して，$P(\{\omega \mid X(\omega) \leq x\})$を$P(X \leq x)$と略記して書くことが多い．本稿では離散型の確率分布の確率質量関数を$f_X(x) = P(\{\omega \mid X(\omega) = x\})$または（同じことであるが）$f_X(x) = P(X = x)$と定義する．また連続型の確率分布に対しては，その分布関数$F_X(x)$を$F_X(x) = P(X \leq x)$と定義し，確率密度関数$f_X(x)$を$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad (-\infty < x < \infty)$を満たす関数として定義する [*13]．

以下では確率変数を$X_1, \ldots, X_T$のように大文字で表記し，しばしばこれらをまとめて$X_{1,T}$と略記する．またこれらに対応する実現値を$x_1, \ldots, x_T$で表記し，同様に，これらをまとめて$x_{1,T}$と略記する．また今後，確率変数K, Λ, Mが登場するが，これらに対応する実現値をk, λ, μのような小文字で表す．

$Y = y$が与えられたXの条件付き確率密度関数を

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

で定義する．ここで，$f_Y(y) = \int f_{X,Y}(x, y)dx$は，$Y = y$における$f_{X,Y}(x, y)$の周辺確率密度関数を表し，積分範囲は定義域すべてを渡るものとする．（$f_X(x)$などについても同様に定義する [*14]）．また本稿で扱う確率分布については，（確率質量関数および確率密度関数に対する）

[*11] 新しく得られた標本を次々ヒストグラムに積み足していく状況を思い描くとイメージがわきやすいかもしれない．たくさんサンプリングされたものは，確率的によく発生するという認識である．

[*12] 本稿では測度論には深く立ち入らない．詳しくは, [10] を見よ．また数理統計学の一般的な内容については, [8] や [12] を見よ．

[*13] $f_X(t)$ の存在は仮定する．

ベイズの定理

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x)\ f_X(x)}{\int f_{Y|X}(y \mid x)\ f_X(x)\ dx}$$

が成り立つことを前提とする．

4.3 ガンマ分布について

後の節に，ガンマ分布が何度も登場する．確率変数Xがガンマ分布$\mathrm{Ga}(\alpha, \beta)$に従うとすると，その確率密度関数 $f_X(x)$ は

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta x}$$

で与えられる[*15]．ここで，αとβはそれぞれ形状パラメータと尺度パラメータとよばれ，$\Gamma(\cdot)$はガンマ関数である．ガンマ分布は，指数分布やカイ二乗分布などを含む分布として知られているが，マルコフ連鎖モンテカルロ法の文脈では，統計モデルとしてポアソン分布を仮定し，事前分布としてガンマ分布を仮定したとき，その事後分布もガンマ分布になるという性質がよく用いられる．

統計モデルとは，雑破にはデータの従う分布であった．今，統計モデルが平均νのポアソン分布に従うとする．このとき事前分布をガンマ分布で与えると，その事後分布はガンマ分布で与えられることを説明しよう．本節で説明する内容を少し複雑にしたもの（2 変数にしたもの）を後で我々は使うことになる．

上でも述べたようにポアソン分布の確率質量関数は次式で与えられる：

$$P(X_t = x_t) = \frac{\nu^{x_t}}{x_t!} e^{-\nu}.$$

ここで，x_t は非負整数，ν は分布の平均である．ポアソン分布に従うデータ $x_{1,T}$に対し，尤度関数[*16] $L(\nu \mid x_{1,T})$ は

$$L(\nu \mid x_{1,T}) = \prod_{t=1}^{T} \frac{\nu^{x_i}}{x_i!} e^{-\nu} \propto \nu^{\sum_{t=1}^{T} x_i} e^{-T\nu}$$

である．$\alpha_1 = \alpha_0 + \sum_{t=1}^{T} x_t, \beta_1 = \beta_0 + T$ とする．ν の事前分布にガンマ分布 $\mathrm{Ga}(\alpha_0, \beta_0)$

$$f(\nu) = \frac{\beta_0^{\alpha_0}}{\Gamma\left(\alpha_0\right)} \nu^{\alpha_0 - 1} e^{-\beta_0 \nu}$$

を使って，νの事後分布 $f(\nu \mid x_{1,T})$を導出すると，νの事後分布はベイズの定理を使って

---

*14 本稿では離散型確率変数の和に対しても，連続型の確率変数と同じ記号（すなわち積分記号）を使って表す．

*15 ガンマ分布の確率密度関数を

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \cdot \frac{x^{\alpha-1}}{\beta^\alpha} \cdot e^{-\frac{x}{\beta}}$$

で与える流儀もある．しかしながら，マルコフ連鎖モンテカルロ法の文脈で使うときは，本文にある定義を採用することが多い．これは，ガンマ分布の確率密度関数（やそれを用いた式）を具体的に記述したときに出てくる（様々な）式の形が，本文の定義のものと比べて煩雑になるからである．

*16 尤度関数を本文のような書き方の代わりに，$L(x_{1,T} \mid \nu)$ のように文字の場所を反対にして書く流儀もある．本稿では何が変数かをわかりやすくするために，変数 ν を先頭に書く流儀を採用した．

$$f(\nu \mid x_{1,T}) \propto L(\nu \mid x_{1,T}) f(\nu) \propto \nu^{\alpha_1 - 1} e^{-\beta_1 \nu}$$

となる．ゆえにこの事後分布はガンマ分布$\mathrm{Ga}(\alpha_1, \beta_1)$であることがわかる．

以上より，νの事前分布がパラメータα_0とβ_0のガンマ分布$\mathrm{Ga}(\alpha_0, \beta_0)$であるとすると，データ$x_{1,T}$が与えられたときの$\nu$の事後分布は，更新されたパラメータ$\alpha_1 = \alpha_0 + \sum_{t=1}^{T} x_t$と$\beta_1 = \beta_0 + T$を持つガンマ分布$\mathrm{Ga}(\alpha_1, \beta_1)$となる．

4.4 マルコフ連鎖と遷移核について

マルコフ連鎖は確率モデルの一種であり，起こりうる事象の連続性を記述するもので，各事象の確率は前の事象で達成された状態にのみ依存するとされる．これはマルコフ連鎖の無記憶性として知られ，将来の行動が現在の状態に至るまでのステップに依存しないことを意味する[*17]．

また遷移核（推移核，または，カーネル）とは，マルコフ連鎖において，現在の状態から次の状態への遷移確率を表す分布である．Y_tをt期の確率変数とし，その確率過程を$(Y_t)_{t=0}^{\infty}$と表すことにする．さらに，Y_tの実現値をy_tとすると，一般的な確率過程においては，確率変数Y_tの確率密度関数は，

$$f(y_t \mid y_0, \ldots, y_{t-1})$$

と過去の$\{Y_{t'}\}_{t'=0}^{t-1}$に依存した条件付き確率密度関数となる．確率変数Y_tの確率密度関数が，1期前の確率変数の実現値のみに依存する，すなわち$\{Y_t\}_{t=0}^{\infty}$の同時確率密度関数が

$$f(y_0, \ldots, y_t) = f(y_0) \cdot f(y_1 \mid y_0) \cdots f(y_t \mid y_{t-1})$$

のように与えられるとき，この確率過程はマルコフ連鎖に従うという．さらにこのとき，$f(y_t \mid y_{t-1})$を通常，遷移核とよぶ．本稿で行うギブスサンプリング法などのマルコフ連鎖モンテカルロ法では，適切な遷移核を設定することが重要である．

5 K, Λ, Mの事後分布について

期間$t = 1, 2, \ldots, T$における日ごとの売買取引の回数$x_1, \ldots, x_T$について，次のように仮定する：与えられた整数k $(1 \leq k \leq T)$に対して，確率変数$X_1, \ldots, X_k$は平均λのポアソン分布$\mathrm{Po}(\lambda)$に従い，確率変数$X_{k+1}, \ldots, X_T$は平均μのポアソン分布$\mathrm{Po}(\mu)$に従う．

5.1 事前分布と初期値について

パラメータK, Λ, Mの事前分布が独立であること，すなわち$f_{K,\Lambda,M}(k, \lambda, \mu) = f_K(k)\, f_\Lambda(\lambda)\, f_M(\mu)$を仮定する．またさらに，事前分布のパラメータは

[*17] マルコフ連鎖は，規約性（有限回のマルコフ連鎖で，任意の状態から（確率密度が正である）任意の状態へ移ることができる性質）と非周期性をもち，不変分布に収束することが一般に知られている．詳しくは [11] を見よ．

$$K \sim \mathrm{U}(1,T), \quad \Lambda \sim \mathrm{Ga}(\alpha_0,\beta_0), \quad M \sim \mathrm{Ga}(\gamma_0,\delta_0) \tag{1}$$

であることを一旦仮定する．この記法は確率変数がどのような分布に従っているかを表すものである．例えば $K \sim \mathrm{U}(1,T)$ という記法は，確率変数Kが離散型一様分布$\mathrm{U}(1,T)$に従うこと，すなわちKは1からTまでの整数の中からランダムに選ばれることを示している．

通常ベイズ統計学では，確率変数などを大文字を使わず，小文字のみを使って表すことが一般的である．その理由は，K,Λ,M のような大文字を使うと，あたかもこれらのパラメータが（本来知り得るはずのない）真の分布に従っているかのような誤解や印象を与えるからである．しかしながら本稿ではそのような一般的な記述法には従わず，その時々で真だと思われる分布に従う確率変数を (1) のように表記することにする．その意図は，パラメータが確率変数であるかどうかをきちんと区別したいがためであり，小文字だけではそれを表現できないからである．

なお上の (1) で，$\alpha_0,\beta_0,\gamma_0,\delta_0$ はハイパー・パラメータと呼ばれ，これらはギブスサンプリング法の各ステップで毎回同じ値が使われる．$\alpha_0,\beta_0,\gamma_0,\delta_0$ には適当な値を恣意的に決めて代入する．今回は0.01 などの小さな正の数を取っておくと問題は起きない．正の数である理由は，ガンマ分布の定義における形状パラメータと尺度パラメータの定義域を確認すればわかる．

最後に，上記の式 (1) に似た記法

$$k \sim \mathrm{U}(1,T), \quad \lambda \sim \mathrm{Ga}(\alpha_0,\beta_0), \quad \mu \sim \mathrm{Ga}(\gamma_0,\delta_0)$$

を紹介して本節を終えよう．この式の意味は，k や λ や μ をその右辺に書いてある分布からサンプリングすることを表しており，ベイズ統計学でしばしば使用される．我々は本稿で行うギブスサンプリングにおいて，k,λ,μ の初期値 $k^{(0)},\lambda^{(0)},\mu^{(0)}$ を決めておく必要がある．これらは上の事前分布からランダムに取得してもよいし，自分で恣意的に決めてもよい [*18]．本稿では，$k^{(0)}$ を全期間の内の真ん中（すなわち $T/2$ を四捨五入したもの）とし，$\lambda^{(0)},\mu^{(0)}$ をそれぞれ，前半のデータ $x_{1,k}$ の平均値，後半のデータ $x_{k+1,T}$ の平均値とする．

5.2 統計モデルとそれに対応する尤度関数

統計モデルとは，サンプルデータの生成に関する一連の統計的仮定を具体化した数学的モデル (S,P) である．ここでSは標本空間（観測可能な値の集合），PはS上の確率分布の集合であり [*19]，雑破には，これまで何度か述べたように「データの従う分布を統計モデルという」と思っておけばよい．

さて，現在の我々の状況において，統計モデルを考えよう．日ごとの売買取引の回数に対応する確率変数を $X_1,\ldots,X_T$ とし，先の3節での考察をもとに，我々はそれらがそれぞれポアソン分布に従うことを仮定する．より正確に述べると，

[*18] もちろん，目標分布への収束の速さへの影響があるので，適切でありそうな値を取っておく方が望ましい．
[*19] S上の確率分布の集合Pは，標本空間S（観測可能な値の集合）に対する，すべての可能な確率分布の集まりを指すものとする．

$$X_t \sim \begin{cases} \mathrm{Po}(\lambda) & t = 1, \ldots, k, \\ \mathrm{Po}(\mu) & t = k+1, \ldots, T \end{cases}$$

を仮定する．ここで，条件付き確率

$$P(X_{1,T} = x_{1,T} \mid K = k, \Lambda = \lambda, M = \mu) = \left(\prod_{t=1}^{k} \frac{\lambda^{x_t}}{x_t!} e^{-\lambda} \right) \cdot \left(\prod_{t=k+1}^{T} \frac{\mu^{x_t}}{x_t!} e^{-\mu} \right)$$

を考えると，尤度関数は観測値（データ） $x_{1,T}$ に対する k, λ, μ の関数として，

$$\begin{aligned} L(k, \lambda, \mu, \mid x_{1,T}) &= \left(\prod_{t=1}^{k} \frac{\lambda^{x_t}}{x_t!} e^{-\lambda} \right) \cdot \left(\prod_{t=k+1}^{T} \frac{\mu^{x_t}}{x_t!} e^{-\mu} \right) \\ &\propto \lambda^{\sum_{t=1}^{k} x_t} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t} e^{-(T-k)\mu} \end{aligned} \tag{2}$$

と定義できる [*20][*21]．

5.3 K, Λ, M の同時事後分布について

構造変化を伴うポアソン分布の同時事後分布について説明する．同時事後分布は，観測されたデータが与えられたときの複数のパラメータが結合した確率分布 $f_{K,\Lambda,M|X_{1,T}}(k, \lambda, \mu \mid x_{1,T})$ である．ベイズ統計学では，観測データを考慮した後のパラメータの値に関する我々の更新された信念（事後分布）を表す．我々が今考えている状況では，観察されたデータが与えられたときの K, Λ, M の同時事後分布に興味がある．この分布は，観測された日ごとの売買取引の回数を考慮した後，これらのパラメータの値に関する事後分布（データを用いて更新された我々の新たな信念とよぶべきもの）を表している．

今，事前分布が独立であること，および仮定 (1) から

$$\begin{aligned} f_{K,\Lambda,M}(k, \lambda, \mu) &= f_K(k)\, f_\Lambda(\lambda)\, f_M(\mu) \\ &= \frac{1}{T} \cdot \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda} \cdot \frac{\delta_0^{\gamma_0}}{\Gamma(\gamma_0)} \mu^{\gamma_0 - 1} e^{-\delta_0 \mu} \end{aligned}$$

となる．したがってベイズの定理と (2) から， K, Λ, M の同時事後分布は

$$\begin{aligned} & f_{K,\Lambda,M|X_{1,T}}(k, \lambda, \mu \mid x_{1,T}) \\ & \propto f_{K,\Lambda,M}(k, \lambda, \mu) \cdot L(k, \lambda, \mu \mid x_{1,T}) \\ & \propto \lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda} \cdot \mu^{\gamma_0 - 1} e^{-\delta_0 \mu} \cdot \lambda^{\sum_{t=1}^{k} x_t} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t} e^{-(T-k)\mu} \\ & = \lambda^{\alpha_0 - 1 + \sum_{t=1}^{k} x_t} e^{-(k+\beta_0)\lambda} \cdot \mu^{\gamma_0 - 1 + \sum_{t=k+1}^{T} x_t} e^{-(T-k+\delta_0)\mu} \end{aligned}$$

となる．上の式を見てわかるように，同時事後分布の正確な形は，尤度関数と事前分布の具体

[*20] 記号$\propto$は，比例を表している．

[*21] 尤度関数は確率ではないので，特定の変数における領域で和や積分が，必ずしも 1 になる必要はないことに注意する．本稿での尤度関数は上記のように定義したが，$P(K = k, \Lambda = \lambda, M = \mu \mid X_{1,T} = x_{1,T})$ に比例する関数からなる同値類（の元のいずれか）と思って定義してもよい．

的な形に依存する．k の事前分布は，変化点がいつ発生したかという信念（事前分布）に基づいて選ぶことができるが，ここでは特に構造変化点 k に対する事前の信念と呼ぶべきものは無いと思って，それを一様分布であると思うことにする．また我々の考えている状況では，ポアソン分布を元に尤度関数を定めているため，λ と μ に事前分布としては，4.3 節で説明したガンマ分布を使用する．

5.4 K の事後分布について

我々が考察しているモデルにおいてのギブスサンプリング法では，現在の λ, μ の値と観測データ $x_{1,T}$ があれば，以下で与える完全条件付き分布から k をサンプリングすることができる．この完全条件付き分布は，k の取り得る値（1から T まで，T は全日数）に対する離散型分布である．

事後分布おいて $K = k$ となる確率は，平均 λ のポアソン分布の下でのその時点までのデータの尤度と，平均 μ のポアソン分布の下でのその時点以降のデータの尤度の積に比例する．これを明示的に表すと (2) より

$$f_{K|\Lambda,M,X_{1,T}}(k \mid \lambda, \mu, x_{1,T}) \propto \lambda^{\sum_{t=1}^{k} x_t} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t} e^{-(T-k)\mu}$$

となる．ただしここでの λ と μ は，この時点までに更新された λ と μ の最新の値である．この分布からサンプリングするには，各々の $K = k$ について（すなわち $K = 1, 2, \ldots, T$ について），これらの確率を計算し，それらの合計が 1 になるように正規化し，そこからサンプリングすればよい．すなわち結論は，

$$k^{\text{new}} \sim \frac{\lambda^{\sum_{t=1}^{k} x_t} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t} e^{-(T-k)\mu}}{\sum_{k=1}^{T} \lambda^{\sum_{t=1}^{k} x_t} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t} e^{-(T-k)\mu}}$$

とすればよい．

5.5 Λ, M の事後分布について

我々はデータが従う分布，すなわち，統計モデルをポアソン分布であることを仮定した．このとき，Λ, M の事前分布としてガンマ分布を仮定した場合の Λ, M の事後分布がガンマ分布になることが 4.3 節と同様に示せる．このことについて以下に述べよう．

Λ, M の事前分布をそれぞれ $\mathrm{Ga}(\alpha_0, \beta_0)$，$\mathrm{Ga}(\gamma_0, \delta_0)$ とする．このとき Λ, M が従う事後分布を明示的に表すと (2) より，Λ, M が独立であると思って，それぞれ

$$\begin{aligned} f_{\Lambda|K,X_{1,T}}(\lambda \mid k, x_{1,T}) &\propto \lambda^{\alpha_0 - 1 + \sum_{t=1}^{k} x_t} e^{-(k+\beta_0)\lambda}, \\ f_{M|K,X_{1,T}}(\mu \mid k, x_{1,T}) &\propto \mu^{\gamma_0 - 1 + \sum_{t=k+1}^{T} x_t} e^{-(T-k+\delta_0)\mu} \end{aligned} \tag{3}$$

となる．ただし，ここでの λ, μ は，この時点までに更新された λ と μ の最新の値である．以上から Λ の事前分布を $\mathrm{Ga}(\alpha_0, \beta_0)$，$M$ の事前分布を $\mathrm{Ga}(\gamma_0, \delta_0)$ と定めると，事後分布はそれぞれ $\mathrm{Ga}(\alpha_1, \beta_1)$，$\mathrm{Ga}(\gamma_1, \delta_1)$ となることがわかった．したがって，

$$\lambda^{\text{new}} \sim \text{Ga}\left(\alpha_0 + \sum_{t=1}^{k} x_t, \beta_0 + k\right),$$
$$\mu^{\text{new}} \sim \text{Ga}\left(\gamma_0 + \sum_{t=k+1}^{T} x_t, \delta_0 + T - k\right)$$

とすればよい.

上記に基づけば，後に第 7 節で述べるような実装が可能である．最後に実装にあたって，重要な注意を述べよう．上の式 (3) を見ると，右辺の式の中に $\lambda^{\sum_{t=1}^{k} x_t}$ がある．ここで，λ は 1 期から k 期までの期間の各々の期における売買取引の回数の平均と概ね思え，各 x_t は実際の売買高のデータであった．もし仮に，各 x_t が 3,000 程度の大きな数であった場合，λ も 3,000 程度の大きな数になることが想定される．その場合，例えば仮に $k = 1{,}000$として，上の式 (3) には，$3000^{3000\times 1000} \fallingdotseq 5.81 \times 10^{10431363}$ ような数が代入されることになる．もちろん理論上はこのように非常に大きな数を用いて計算することも可能であるが，実際の計算機ではこれを扱うのが難しい場合もあるので，指数関数が分布に含まれている場合は注意が必要である．プログラムを組むときに，対数関数を使うなどの工夫をすればこの点を緩和できるが[*22]，その方法も完全ではないため，どのような大きさの数字のデータを扱うのかには，注意をしておく必要がある.

6 分布の収束性

ギブスサンプリング法の収束性は , アルゴリズムが定常分布に収束するかどうかを決定する重要な指標である．収束性が保証されていない場合，アルゴリズムが定常分布に収束しない可能性があり，その結果，サンプリングされた値が意味をもたない可能性がある．そのため，ギブスサンプリング法の収束性の証明は，アルゴリズムの信頼性を確保する上で重要である．本節においては，その収束性について議論する.

6.1 詳細釣り合い条件の成立と事後分布の収束性

遷移確率とは，ある状態から別の状態に移る確率のことであり，対象とする確率分布や事後確率などに基づいて定義される．パラメータ k, λ, μ の更新順序をランダムに等確率で定めるとき，遷移確率 $P((k,\lambda,\mu) \to (k',\lambda',\mu'))$ は次のように表される.

$$\begin{aligned}
&P((k,\lambda,\mu) \to (k',\lambda',\mu')) \\
&= \frac{1}{6} f_{K|\Lambda,M}(k' \mid \lambda,\mu)\ f_{\Lambda|K,M}(\lambda' \mid k',\mu)\ f_{M|K,\Lambda}(\mu' \mid k',\lambda') \\
&\quad + \frac{1}{6} f_{K|\Lambda,M}(k' \mid \lambda,\mu)\ f_{M|K,\Lambda}(\mu' \mid k',\lambda)\ f_{\Lambda|K,M}(\lambda' \mid k',\mu') \\
&\quad + \frac{1}{6} f_{\Lambda|K,M}(\lambda' \mid k,\mu)\ f_{K|\Lambda,M}(k' \mid \lambda',\mu)\ f_{M|K,\Lambda}(\mu' \mid k',\lambda')
\end{aligned}$$

[*22] 例えば $\log 1000 \fallingdotseq 6.91$のように，対数をとると扱う数字を小さくすることができる.

$$
\begin{aligned}
&+\frac{1}{6} f_{\Lambda|K,M}(\lambda' \mid k,\mu)\, f_{M|K,\Lambda}(\mu' \mid k,\lambda')\, f_{K|\Lambda,M}(k' \mid \lambda',\mu') \\
&+\frac{1}{6} f_{M|K,\Lambda}(\mu' \mid k,\lambda)\, f_{K|\Lambda,M}(k' \mid \lambda,\mu')\, f_{\Lambda|K,M}(\lambda' \mid k',\mu') \\
&+\frac{1}{6} f_{M|K,\Lambda}(\mu' \mid k,\lambda)\, f_{\Lambda|K,M}(\lambda' \mid k,\mu')\, f_{K|\Lambda,M}(k' \mid \lambda',\mu').
\end{aligned}
$$

上記はパラメータの更新の順番に対応しており，右辺は合計 $3! = 6$ 項ある．右辺の最初にある $f_{K|\Lambda,M}(k' \mid \lambda,\mu)\, f_{\Lambda|K,M}(\lambda' \mid k',\mu)\, f_{M|K,\Lambda}(\mu' \mid k',\lambda')$ は，$k \to k', \lambda \to \lambda', \mu \to \mu'$ の順で更新されること，より丁寧に述べると，

$$
\begin{pmatrix} k \\ \lambda \\ \mu \end{pmatrix} \to \begin{pmatrix} k' \\ \lambda \\ \mu \end{pmatrix} \to \begin{pmatrix} k' \\ \lambda' \\ \mu \end{pmatrix} \to \begin{pmatrix} k' \\ \lambda' \\ \mu' \end{pmatrix}
$$

のように更新されることを表している．（上記の式では，各項が異なる変数の更新順序を表している．）

このとき，

$$
\begin{aligned}
&P((k,\lambda,\mu) \to (k',\lambda',\mu')) \\
&= \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k',\lambda,\mu)}{f_{\Lambda,M}(\lambda,\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu)}{f_{K,M}(k',\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu')}{f_{K,\Lambda}(k',\lambda')} \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k',\lambda,\mu)}{f_{\Lambda,M}(\lambda,\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda,\mu')}{f_{K,\Lambda}(k',\lambda)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu')}{f_{K,M}(k',\mu')} \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda',\mu)}{f_{K,M}(k,\mu)} \cdot \frac{f_{K,M,\Lambda}(k',\lambda',\mu)}{f_{\Lambda,M}(\lambda',\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu')}{f_{K,\Lambda}(k',\lambda')} \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda',\mu)}{f_{K,M}(k,\mu)} \cdot \frac{f_{K,\Lambda,M}(k,\lambda',\mu')}{f_{K,\Lambda}(k,\lambda')} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu')}{f_{\Lambda,M}(\lambda',\mu')} \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda,\mu')}{f_{K,\Lambda}(k,\lambda)} \cdot \frac{f_{K,M,\Lambda}(k',\lambda,\mu')}{f_{\Lambda,M}(\lambda,\mu')} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu')}{f_{K,M}(k',\mu')} \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda,\mu')}{f_{K,\Lambda}(k,\lambda)} \cdot \frac{f_{K,\Lambda,M}(k,\lambda',\mu')}{f_{K,M}(k,\mu')} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu')}{f_{\Lambda,M}(\lambda',\mu')}
\end{aligned}
$$

だから，

$$
\begin{aligned}
&f_{K,\Lambda,M}(k,\lambda,\mu) P((k,\lambda,\mu) \to (k',\lambda',\mu')) \\
&= \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda,\mu)}{f_{\Lambda,M}(\lambda,\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda,\mu)}{f_{K,M}(k',\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu)}{f_{K,\Lambda}(k',\lambda')} \cdot f_{K,\Lambda,M}(k',\lambda',\mu') \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda,\mu)}{f_{\Lambda,M}(\lambda,\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda,\mu)}{f_{K,\Lambda}(k',\lambda)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda,\mu')}{f_{K,M}(k',\mu')} \cdot f_{K,\Lambda,M}(k',\lambda',\mu') \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda,\mu)}{f_{K,M}(k,\mu)} \cdot \frac{f_{K,M,\Lambda}(k,\lambda',\mu)}{f_{\Lambda,M}(\lambda',\mu)} \cdot \frac{f_{K,\Lambda,M}(k',\lambda',\mu)}{f_{K,\Lambda}(k',\lambda')} \cdot f_{K,\Lambda,M}(k',\lambda',\mu') \\
&\quad + \frac{1}{6} \cdot \frac{f_{K,\Lambda,M}(k,\lambda,\mu)}{f_{K,M}(k,\mu)} \cdot \frac{f_{K,\Lambda,M}(k,\lambda',\mu)}{f_{K,\Lambda}(k,\lambda')} \cdot \frac{f_{K,\Lambda,M}(k,\lambda',\mu')}{f_{\Lambda,M}(\lambda',\mu')} \cdot f_{K,\Lambda,M}(k',\lambda',\mu')
\end{aligned}
$$

$$+\frac{1}{6}\cdot\frac{f_{K,\Lambda,M}(k,\lambda,\mu)}{f_{K,\Lambda}(k,\lambda)}\cdot\frac{f_{K,M,\Lambda}(k,\lambda,\mu')}{f_{\Lambda,M}(\lambda,\mu')}\cdot\frac{f_{K,\Lambda,M}(k',\lambda,\mu')}{f_{K,M}(k',\mu')}\cdot f_{K,\Lambda,M}(k',\lambda',\mu')$$
$$+\frac{1}{6}\cdot\frac{f_{K,\Lambda,M}(k,\lambda,\mu)}{f_{K,\Lambda}(k,\lambda)}\cdot\frac{f_{K,\Lambda,M}(k,\lambda,\mu')}{f_{K,M}(k,\mu')}\cdot\frac{f_{K,\Lambda,M}(k,\lambda',\mu')}{f_{\Lambda,M}(\lambda',\mu')}\cdot f_{K,\Lambda,M}(k',\lambda',\mu')$$
$$= f_{K,\Lambda,M}(k',\lambda',\mu')\ P((k',\lambda',\mu')\to(k,\lambda,\mu)).$$

となるため，詳細釣合い条件と呼ばれる，次の式が成り立つことがわかる．

$$f_{K,\Lambda,M}(k,\lambda,\mu)\ P((k,\lambda,\mu)\to(k',\lambda',\mu')) = f_{K,\Lambda,M}(k',\lambda',\mu')\ P((k',\lambda',\mu')\to(k,\lambda,\mu)).$$

ここで，上の式の両辺において，すべての (k,λ,μ) について和をとると，

$$\int f_{K,\Lambda,M}(k,\lambda,\mu)\ P((k,\lambda,\mu)\to(k',\lambda',\mu'))\ dk\ d\lambda\ d\mu$$
$$= f_{K,\Lambda,M}(k',\lambda',\mu')\ \int P((k',\lambda',\mu')\to(k,\lambda,\mu))\ dk\ d\lambda\ d\mu$$
$$= f_{K,\Lambda,M}(k',\lambda',\mu')$$

が得られ，定常分布に収束することがわかる．

6.2 収束診断について

前節で述べた事柄によって，我々の方法で分析を行えば，所与の仮定の下で無限回のサンプリングを行えば，分布が収束していることがわかった．しかしながら，有限の時間の実行で，どれくらい収束するかの確認（収束診断）は重要である．収束診断の方法は様々あることが知られているが，本稿では，広く使われていて実装が簡単な Geweke の収束診断 [1,4] を用いる．

Geweke の収束診断は，マルコフ連鎖モンテカルロ法によるサンプリングで収束性を評価するために使われる方法の 1 つであるが，この方法は標本列の（Burn-in 期間を除いた）はじめと終わりの部分とを比較して，それらの平均が異なるかどうかを検定することによって，標本列が収束したかどうかを判断する方法である．より具体的には，標本列の最初の 10% と最後の 50% を取り出して，それらの 2 つの平均が統計的に異なるかどうかを次の統計量 Z による Z 検定によって判断する．

$$Z = \frac{\bar{X_1} - \bar{X_2}}{\sqrt{V_1(\bar{X_1}) + V_2(\bar{X_2})}}.$$

ここで $V_1(\bar{X_1})$ と $V_2(\bar{X_2})$ はそれぞれ $\bar{X_1}$ と $\bar{X_2}$ の分散の推定量である [*23]．Z 値が正規分布の 95% 信頼区間（− 1.96 から 1.96）の外側にある場合，両者は統計的に異なるとみなされ，その場合，標本は収束していないとみなされる． Geweke の収束診断については，プログラミング言語 R を用いた場合は，パッケージが用意されているので，該当箇所を数行書き加えるだ

[*23] ここでは，その推定量の正確な求め方については割愛する．興味のある読者は，[1] を参照されたい．

けで，収束診断ができる．（該当箇所は，プログラムの説明の項で示す．R 以外のプログラミング言語を用いる場合は，[4] を参考に，比較的容易にプログラムが組める．）

収束診断について，一般的には，複数の方法で収束性を判定するべきであるとされているが，本稿では Geweke の収束診断のみを用いることにする．

7　株式市場における投資家の売買取引への適用例

ここでは，ギブスサンプリング法を用いた構造変化に関する分析について，株式市場における投資家の売買取引への適用例を示す．すでに述べたように，対象企業はANA Holdings，対象期間は 2018 年 1 月 1 日から 2023 年 6 月 23 日までの期間である．分析にはフリー統計ソフト R を用いており，データは Refinitiv Eikon から取得している [*24]．

なお，ここでの分析では，利用可能なデータの制約により，実際の売買取引の回数に関するデータではなく，売買高（出来高）[*25] を用いていることに留意してほしい．売買高とは，ある期間（例：1 取引日）において売買された株式数（例：1,000,000 株など）のことを意味する．銘柄ごとに，最小の取引単位（1 単元）が決められている．この取引単位は単元株数と呼ばれ，ANA Holdings は 1 単元 100 株である．売買高 1,000,000 株の場合，すべての売買取引が 1 単元で行われているとすれば，1 日の売買取引の回数は 10,000 回であるが，10 単元などの単位で売買取引が行われることもあるため，売買高のデータから厳密な売買取引の回数を知ることはできない．また，本稿で用いるデータの数値は大きいため，ポアソン分布の持つ良い性質を若干犠牲にしていることにも注意されたい．

7.1　データの概要

本稿の分析で用いる売買高のデータは Refinitiv Eikon から取得しているが，そのデータは表 1 のようなレイアウトになっている．ここで，1 行目 1 列目のセルに「Name」，1 行目 2 列目のセルに「ANA HOLDINGS - TURNOVER BY VOLUME」という文字列が「標準」の表示形式で入っており，2 行目 1 列目のセルに日付のデータが「日付」の表示形式で文字列が入っている．また，2 行目 2 列目のセルに売買高の値が「数値」の表示形式で文字列が入っている．単位は千株であるため，「1577.9」は売買高が 1,577,900 株であったことを意味する．売買取引が行われていない場合，2 行目 2 列目のセルに売買高の値は「NA」になっている．以下，同様のデータが 2023 年 6 月 23 日まで入っている．

[*24] Refinitiv Eikon とは，金融市場の専門家向けに，財務・金融を中心とするデータを提供するツールのことである．Excel のアドインを利用し，各種のデータを容易にダウンロードすることができる．データベースの詳細については，こちら（https://www.refinitiv.com/content/dam/marketing/en_us/documents/brochures/eikon-overview-brochure.pdf）を参照してほしい．

[*25] https://www.jpx.co.jp/glossary/ha/359.html

表 1　売買高のデータ

Name	ANA HOLDINGS - TURNOVER BY VOLUME
2018/1/1	NA
2018/1/2	NA
2018/1/3	NA
2018/1/4	1577.9
2018/1/5	1662.2
2018/1/9	1721.1

7.2 分析のプログラム

分析のプログラムは下記の通りである．1 行目から 19 行目では，分析に必要な R のライブラリと Refinitiv Eikon からダウンロードしたデータを読み込み，分析しやすいようにデータを整形している．売買高の値が「NA」になっている行は削除し，売買高の数値がある日のみを残している．なお，17 行目で売買高のデータを 1,000 で割っているのは，5.5 節で説明した理由による．

21 行目から 50 行目では，全期間を表す変数 T や，ハイパーパラメータの設定などを行っている．特に，パラメータの初期値，反復回数， burn-in 期間の設定は重要になる．構造変化が起きた時点を表す k については，全期間の日付の中央値，構造変化前の日ごとの売買取引の回数の平均を表す λ は，k 以前の期間における売買高の平均値，構造変化後の日ごとの売買取引の回数の平均を表す μ は，k の翌日以後の期間における売買高の平均値としている．

52 行目から 105 行目では，ギブスサンプリング法による k, λ, μ のサンプリングを行っている．本稿では，k, λ, μ のサンプリングを行う際に，それぞれのフェーズごとに，サンプリングの順番をランダムに決定している．反復回数 20,000 回， burn-in 期間 1,000 回であるため，1,001 回目のサンプリングから，サンプリングされた k, λ, μ を保存することになる．107 行目から 110 行目において，収束診断を行っている．

112 行目以降については，分析の結果を Excel に出力するための処理である．

```
1  # ライブラリの読み込み
2  library(readxl)
3  library(coda)
4
5  # すべての変数のデータを一旦削除する
6  rm(list = ls())
7
8  # Refinitiv Eikonからダウンロードしたデータの読み込み
9  anatradvol <- read_excel("anatradvol2018010120230623.xlsx")
10
```

```
# 変数名を変更する
names(anatradvol) <- c("date", "tradvol")

# 売買高がNAの行を削除する
anatradvol <- anatradvol[!(anatradvol$tradvol=="NA"),]

# 売買高のデータを1000で割る
anatradvol$tradvol <- as.numeric(anatradvol$tradvol)
anatradvol$tradvol <- anatradvol$tradvol/1000

# 全期間を表す変数T(total_period)を定義する
total_period <- nrow(anatradvol)
print(total_period)

# ハイパーパラメータの設定
lambda_shape_hyperparameter <- 0.01
lambda_rate_hyperparameter <- 0.01
mu_shape_hyperparameter <- 0.01
mu_rate_hyperparameter <- 0.01

# パラメータの初期値
# kの初期値を設定する（日付の中央値）
k <- round(as.numeric(median(anatradvol$date)),0)
# lambdaの初期値を設定する（kの初期値までの期間のtradvolの平均値）
lambda <- mean(anatradvol$tradvol[anatradvol$date <= k])
print(paste("lambda =", lambda))
# muの初期値設定する（kの初期値よりも後の期間のtradvolの平均値）
mu <- mean(anatradvol$tradvol[anatradvol$date > k])
print(paste("mu =", mu))

# 反復回数
n_iter <- 100000

# burn-in期間
burn_in <- 1000

# パラメーター値を格納するベクトル
k_samples <- numeric(n_iter)
```

```
lambda_samples <- numeric(n_iter)
mu_samples <- numeric(n_iter)

# ギブスサンプリング（k, lambda, muの順番ではなく，ランダムにサンプリング）
for (i in 1:(n_iter)) {
  random_order <- sample(1:3, 3, replace = FALSE) # 一様分布からランダムな数値の組み合わせを生成する
  random_order <- as.list(random_order)
  # 1: kのサンプリング
  # 2: lambdaのサンプリング
  # 3: muのサンプリング
  for (j in 1:3) {
    if (random_order[[j]] == 1) {
      # kのサンプリング
      prob_k <- numeric(total_period)
      for (t in 1:total_period) {
        #print(paste("t =", t))
        if (t < total_period) {
          # log変換を使用して数値的な安定性を確保
          log_prob_k <- sum(anatradvol$tradvol[1:t])*log(lambda) + (sum(anatradvol$tradvol[1:total_period])-sum(anatradvol$tradvol[1:t]))*log(mu) - lambda*t - mu*(total_period-t)
          #print(paste("log_prob_k =", log_prob_k))
          prob_k[t] <- exp(log_prob_k)
          #print(paste("prob_k =", prob_k[t]))
        } else if (t == total_period) {
          # log変換を使用して数値的な安定性を確保
          log_prob_k <- sum(anatradvol$tradvol[1:t])*log(lambda) - lambda*t
          prob_k[t] <- exp(log_prob_k)
          #print(paste("prob_k[T] =", prob_k[t]))
        }
      }

      prob_k <- prob_k / sum(prob_k, na.rm = TRUE)
      #print(paste("全確率=", sum(prob_k)))

      k <- sample(as.numeric(anatradvol$date), size = 1, prob = prob_k)
      print(k)
```

```
      } else if (random_order[[j]] == 2) {
      # lambdaのサンプリング
      lambda_shape_parameter <-sum(anatradvol$tradvol[anatradvol$date
          <= k]) + lambda_shape_hyperparameter
      lambda_rate_parameter <- length(anatradvol$date[anatradvol$date
          <= k]) + lambda_rate_hyperparameter
      lambda <- rgamma(1, shape = lambda_shape_parameter, rate = lambda
          _rate_parameter)
      #print(paste("lambda =", lambda))
      } else if (random_order[[j]] == 3) {
      # muのサンプリング
      mu_shape_parameter <- sum(anatradvol$tradvol[anatradvol$date > k])
          + mu_shape_hyperparameter
      mu_rate_parameter <- length(anatradvol$date[anatradvol$date > k])
          + mu_rate_hyperparameter
      mu <- rgamma(1, shape = mu_shape_parameter, rate = mu_rate_
          parameter)
      #print(paste("mu =", mu))
    }

  # バーンイン期間が終了したら，サンプルを保存する
  if (i > burn_in) {
    k_samples[i - burn_in] <- k
    lambda_samples[i - burn_in] <- lambda
    mu_samples[i - burn_in] <- mu
    }
  }
}

# Gewekeの収束診断
geweke.diag(as.mcmc(k_samples))
geweke.diag(as.mcmc(lambda_samples))
geweke.diag(as.mcmc(mu_samples))

#kの値の表示形式を日付に変更する
turning_k <- as.POSIXct(k, origin = "1970-01-01")
print(turning_k)

```

```
#install.packages("writexl") #一度だけ実行する必要がある
library(writexl)

# 結果データフレームの作成
result_data <- data.frame(
  turning_k = turning_k,
  lambda = lambda,
  mu = mu,
  geweke_k = geweke.diag(as.mcmc(k_samples))$z,
  geweke_lambda = geweke.diag(as.mcmc(lambda_samples))$z,
  geweke_mu = geweke.diag(as.mcmc(mu_samples))$z
)

# 現在の日時を取得
current_time <- Sys.time()

# 日時を文字列に変換
formatted_time <- format(current_time, "%Y-%m-%d-%H-%M-%S")

# ファイル名を作成
file_name <- paste0("result-", formatted_time, ".xlsx")

# 結果データフレームをExcelファイルに保存
write_xlsx(result_data, path = file_name)
```

7.3 分析の結果

分析の結果は表 2 の通りである．ここでは，下記の結果に基づいて，構造変化の有無とその時点について考察する．まず，収束診断の結果を確認する．統計量 Z について，geweke k，geweke λ，geweke μ のいずれも，正規分布の 95% 信頼区間（-1.96 から 1.96）の範囲内にある．このことから，k, λ, μ の標本列は収束したと判断する．

λ は構造変化前における日ごとの売買取引の回数の平均，μ は構造変化後のそれを表す．λ が 1.045，μ が 3.860 であることから，構造変化後に日ごとの売買取引の回数は増加しているといえる．構造変化が起きたと思われる時点が turning k であり，2020 年 2 月 21 日となっている [*26]．この時点については，世界保健機関（WHO）が新型コロナウイルスについて「国際的に懸念される公衆衛生上の緊急事態」（以下，緊急事態宣言）を宣言した 2020 年 1 月 30 日 [3] よりも後の期間である．こうした外生的な要因によって構造変化が起こっているというのが，ありうる 1 つの解釈である．ただし，緊急事態宣言が出された 2020 年 1 月 30 日と分析

表 2　分析の結果

turning k	λ	μ	geweke k	geweke λ	geweke μ
2020-02-21	1.045	3.860	0.715	0.740	0.727

の結果が示している構造変化の時点である 2020 年 2 月 21 日については，若干の日付のラグがある．この点に関して，2020 年 1 月 30 日よりも後の期間に発生している事象を整理する．具体的には，東証株価指数（TOPIX）の推移，ANA Holdingsの適時開示書類 [*27]，日本経済新聞のANA Holdings関連の記事を確認する．TOPIX の推移 [*28] について，2020 年 1 月 30 日から 2020 年 2 月 21 日はほぼ同水準で推移しているが，その後，週明けの最初の取引日である 2020 年 2 月 25 日から下落し，2020 年 3 月 13 日には 1,261.70 となっている．ANA Holdingsの適時開示書類 [*29] について，2020 年 1 月 30 日に第 3 四半期の決算短信，2020 年 2 月 25 日に固定資産（航空機）の取得計画，2020 年 3 月 13 日に代表取締役の異動が開示されている．これらの事象が新型コロナウイルスの感染拡大に起因しているか否かについては，開示された情報からは断言できない．日本経済新聞のANA Holdings関連の記事 [*30] について，2020 年 1 月 30 日から 2020 年 3 月 13 日の期間においては，中国便の運休を皮切りに，国内線の減便も相次いで実施されている．以上より，上記の結果については，次のようにまとめることができる．

1. パンデミックの前後で，ANA Holdingsの日ごとの売買取引の回数は増加しており，構造変化は起きている．
2. 構造変化が起きている時点は 2020 年 2 月 21 日であり，これは，緊急事態宣言が出された 2020 年 1 月 30 日よりも後の時点である．2020 年 1 月 30 日よりも後の期間における TOPIX の動向から，2020 年 2 月 21 日の直後の期間は，市場全体として株価が低迷している時期であるといえる．航空業界に固有の要因としては，中国便の運休，国内線の減便といったものが挙げられる．以上より，2020 年 2 月 21 日における構造変化は，新型コロナウイルスの感染拡大に起因していると判断することが妥当である．

7.4 注意

1. ハイパーパラメータは，プログラムの振る舞いを制御する変数である．これらは一般に

[*26] 同様の分析を複数回繰り返したが，turning k は 2020 年 2 月 20 日，2020 年 2 月 21 日，2020 年 2 月 25 日のいずれかになることが多かった．そのうち，2020 年 2 月 21 日になることが最も多かったため，本稿では，構造変化点を 2020 年 2 月 21 日として結果を考察する．なお，構造変化点に関わらず，λ と μ は表 2 と同様の傾向を示しているため，分析結果の考察に大きな影響はないと考えている．

[*27] 適時開示書類とは，金融商品取引所における適時開示制度のもとで開示される会社情報のことである．詳細は，日本取引所グループの HP（https://www.jpx.co.jp/equities/listing/disclosure/overview/index.html）を参照してほしい．

[*28] https://finance.yahoo.co.jp/quote/998405.T/history?from=20200130&to=20200331&timeFrame=d

[*29] https://www.ana.co.jp/group/investors/irdata/disclosure/

[*30] 日本経済新聞のウェブ版（https://www.nikkei.com/）を対象に，2020 年 1 月 30 日から 2020 年 3 月 13 日の期間について，「ANA，運休，減便」のキーワードで検索した．

プログラムのループが回る前に設定され，そのプロセスの途中で変更されることはない．

2. 3 つのパラメータ λ, μ, k の初期値は，データの平均値を使用している．これらの初期値は主観的に設定することも可能であるが，本稿ではデータに基づいて設定されていることに注意する．
3. エクセルでは，日付データは「シリアル値」と呼ばれる特定の数値を使って管理される．シリアル値とは，「1900 年 1 月 1 日」を「1」としてから何日経過したかを示す数値である．Excel には日付データを操作するための様々な関数が用意されている．例えば，‘DATE’ 関数は年，月，日の 3 つの独立した値を受け取ることができ，これらを組み合わせて日付を作成することができる．さらに，Excel では日付データに「1」を追加して「1 日」，つまり翌日の日付を表示することができる．

8 目標分布を直接得る方法

これまで，ギブスサンプリング法を説明するため，若干込み入った方法を採ってきたが，ここではもっと直接的に目標分布を得る方法を紹介しよう [*31]．この方法は今回の考察対象においては，これまでに説明してきたギブスサンプリング法よりもよい方法であると思われるが，分析する対象が複雑になった場合（特に変数の数が増えた場合）は，使えない可能性があるため注意が必要である．しかしながら K, Λ, M の分布は，サンプリングしなくても直接式で書けるため，本来，本節の方法が適用できるのであれば，この方法の方がよい．

上で説明したギブスサンプリング法では $\alpha_0, \beta_0, \gamma_0, \delta_0$ を正の数としたが，以下の計算では $\alpha_0 = 0, \beta_0 = 0, \gamma_0 = 0, \delta_0 = 0$ として問題は起きないのでこれを仮定する．この設定の元で，目標分布を直接求めることを試みよう．まず第 5 節の (2) にある尤度関数の式

$$L(k, \lambda, \mu \mid x_{1,T}) \propto \lambda^{\sum_{t=1}^{k} x_t - 1} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t - 1} e^{-(T-k)\mu}$$

を思い出すと，

$$\begin{aligned} &\sum_{k=1}^{T} \int_0^\infty \int_0^\infty \lambda^{\sum_{t=1}^{k} x_t - 1} e^{-k\lambda} \cdot \mu^{\sum_{t=k+1}^{T} x_t - 1} e^{-(T-k)\mu} \, d\lambda \, d\mu \\ &= \sum_{k=1}^{T} \frac{\Gamma\left(\sum_{t=1}^{k} x_t\right)}{k^{\sum_{t=1}^{k} x_t}} \cdot \frac{\Gamma\left(\sum_{t=k+1}^{T} x_t\right)}{(T-k)^{\sum_{t=k+1}^{T}}} \end{aligned}$$

となる．今，この式を B とし，

$$B_k = \frac{\Gamma\left(\sum_{t=1}^{k} x_t\right)}{k^{\sum_{t=1}^{k} x_t}} \cdot \frac{\Gamma\left(\sum_{t=k+1}^{T} x_t\right)}{(T-k)^{\sum_{t=k+1}^{T}}}$$

とすると，K, Λ, M の分布はそれぞれ，

[*31] 斎藤新悟氏の指摘による．

$$f_K(k) = \frac{B_k}{B},$$

$$f_\Lambda(\lambda) = \frac{1}{B}\sum_{k=1}^{T}\int_0^\infty A d\mu = \frac{1}{B}\sum_{k=1}^{T}\lambda^{\sum_{t=1}^{k} x_t - 1}e^{-k\lambda}\cdot\frac{\Gamma\left(\sum_{t=k+1}^{T} x_t\right)}{(T-k)^{\sum_{t=k+1}^{T} x_t}},$$

$$f_M(\mu) = \frac{1}{B}\sum_{k=1}^{T}\int_0^\infty A d\lambda = \frac{1}{B}\sum_{k=1}^{T}\mu^{\sum_{t=k+1}^{T} x_t - 1}e^{-k\mu}\cdot\frac{\Gamma\left(\sum_{t=1}^{k} x_t\right)}{k^{\sum_{t=1}^{k} x_t}}$$

となる．

上記を元にここでは，数式を簡明に表記してくれる計算機ソフトウェアMathematicaを用いて，計算が容易なKの確率質量関数と期待値を求めてみよう．Mathematicaのプログラミングコードは，以下の通りである．なお以下のdata.xlsx には，1列目の1行目から1334行目に，日付順に売買高のデータが入っているものとする[*32]．

```
(*Excel ファイルのパスとファイル名を指定する．*)
filePath="data.xlsx";

(*Import 関数を使用して Excel データを取り込む．*)
data=Import[filePath,{"Data",1}];

(*1 列目の 1 行目から 1334 行目までのデータを取得する．*)
T=1334;
ls=data[[1;;T,1]];

(*必要な関数を以下で定義する．*)
sum1[k_]:=Total[ls[[1;;k]]];
sum2[k_]:=Total[ls[[k+1;;T]]];

(*K の分布を以下で定義する．*)
fKpre1[k_]:=If[k<T,(Gamma[sum1[k]]*Gamma[sum2[k]])/(k^sum1[k]*(T-k)^sum2[k]),
            Gamma[sum1[k]]/k^sum1[k]];
fKpre2=Sum[fKpre1[k],{k,1,T}];
fK[k_]:=fKpre1[k]/fKpre2;

efk=Sum[k*fK[k],{k,1,T}];
```

詳しい説明は割愛するが，上のプログラムを用いて実際に計算すると，Kの期待値efkは

[*32] 基本的に，前節までで使ったデータと同じものである.5.5 節では売買高のデータを1, 000で割って作ったデータを用いたが，ここではそれをすることなくプログラムを動かすことにする．

小数第一位を四捨五入すると 519 となっており，これは 2020 年 2 月 21 日に対応する．さらに計算機によって，近似的に

$$
\begin{aligned}
&f_K(1) = 4.92 \times 10^{-224}, \qquad f_K(2) = 4.41 \times 10^{-224},\\
&\cdots\cdots\\
&f_K(517) = 0.055, \qquad f_K(518) = 0.170,\\
&f_K(519) = 0.460, \qquad f_K(520) = 0.205,\\
&f_K(521) = 0.093, \qquad f_K(522) = 0.0022,\\
&\cdots\cdots\\
&f_K(1333) = 3.03 \times 10^{-224}, \qquad f_K(1334) = 1.69 \times 10^{-224}
\end{aligned}
$$

となることがわかる [*33]．

以上の結果を踏まえると，これまでの節で述べてきたギブスサンプリング法による近似は，完全に正確なものではないが，ある程度妥当なものであることが確認できる．

9 おわりに

我々は，株式市場における投資家の売買取引の回数に着目して，それらの取引がポアソン分布に従うことを仮定したモデルについて考察した．特に，日ごとの売買取引の回数がパンデミック前後で変化したか否かと，変化した場合の構造変化点について，キブスサンプリング法を用いて考察した．

考察の結果，次のことが明らかになった．第 1 に，パンデミックの前後で，ANA Holdings の日ごとの売買取引の回数は増加しており，構造変化は起きている．第 2 に，構造変化が起きている時点は 2020 年 2 月 21 日であり，これは，緊急事態宣言が出された 2020 年 1 月 30 日よりも後の時点である．2020 年 1 月 30 日よりも後の期間における TOPIX の動向から，2020 年 2 月 21 日の直後の期間は，市場全体として株価が低迷している時期であるといえる．航空業界に固有の要因としては，中国便の運休，国内線の減便といったものが挙げられる．本稿の分析結果は，ANA Holdingsにおいて，新型コロナウイルスの感染拡大に起因し，日ごとの売買取引の回数に構造変化が起こっていることを示唆している．

謝辞

本論文の作成にあたり，大変有益かつ適切なご助言をいただいた，九州大学の斎藤新悟氏と小野塚友一氏，名古屋大学の広瀬稔氏に心から感謝申し上げます．

[*33] $k = 517$ は 2020 年 2 月 19 日，$k = 518$ は 2020 年 2 月 20 日，$k = 520$ は 2020 年 2 月 25 日，$k = 521$ は 2020 年 2 月 26 日に対応している．

参考文献

[1] John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, 1991.

[2] Christian P. Robert. *The Bayesian choice. From decision-theoretic foundations to computational implementation.* Springer Texts Stat. New York, NY: Springer, 2nd ed., 1st paperback ed. edition, 2007.

[3] BBC ニュース. WHO, 新型コロナの緊急事態宣言を終了 脅威は消えずと警告. Available at `https: //www.bbc.com/japanese/65506117`, 2023. Accessed: 2023 年 6 月 29 日.

[4] coda source: R/geweke.r. Available at `https://rdrr.io/cran/coda/src/R/geweke.R`. Accessed: 2023 年 6 月 26 日.

[5] 林田実. ベイズ統計学と mcmc–メトロポリス・ヘイスティングス法の matlab による実現. 北九州市立大学商経論集 = The Review of business and economics, Vol. 46, 3・4, pp. 147–159, 2011.

[6] 伊庭幸人, 種村正美, 大森裕浩, 和合肇, 佐藤整尚, 高橋明彦. マルコフ連鎖モンテカルロ法とその周辺. 統計科学のフロンティア. 岩波書店, 2005.

[7] 各務和彦. ベイズ分析の理論と応用 : R 言語による経済データの分析. ライブラリデータ分析への招待. 新世社 and サイエンス社 (発売), 2022.

[8] 久保川達也. 現代数理統計学の基礎. 共立講座 数学の魅力. 共立出版, 2017.

[9] 中妻照雄. 入門ベイズ統計学. ファイナンス・ライブラリー. 朝倉書店, 2007.

[10] 清水泰隆. 統計学への確率論, その先へ : ゼロからの測度論的理解と漸近理論への架け橋. 内田老鶴圃, 第 2 版, 2021.

[11] 竹居正登. 入門確率過程. 森北出版, 2020.

[12] 竹村彰通. 現代数理統計学. 学術図書出版社, 新装改訂版, 2020.

[13] 渡辺澄夫. ベイズ統計の理論と方法. コロナ社, 2012.