

論文 / 著書情報
Article / Book Information

題目(和文)	販売データの統計的性質の解析と時系列分析及びその応用
Title(English)	
著者(和文)	迫田元
Author(English)	Gen Sakoda
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11582号, 授与年月日:2020年9月25日, 学位の種別:課程博士, 審査員:高安 美佐子,出口 弘,三宅 美博,青西 亨,小野 功
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11582号, Conferred date:2020/9/25, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

博士学位論文

販売データの統計的性質の解析と
時系列分析及びその応用

令和2年度 9月

迫田 元

Gen Sakoda

東京工業大学大学院 情報理工学院
数理・計算科学系 知能情報コース

目次

第1章 緒言.....	3
1.1 背景.....	3
1.2 目的.....	10
第2章 データ.....	13
2.1 販売時点情報（Point-Of-Sales：POS）の概要.....	13
2.2 本研究で用いた POS データ.....	14
第3章 販売データの統計的性質を考慮した非定常時系列分析手法の確立.....	16
3.1 テイラースケーリング則.....	17
3.2 状態空間モデル.....	21
3.3 粒子フィルタとその拡張.....	26
3.4 効果検証.....	33
3.5 まとめ.....	40
第4章 非定常時系列分析手法の廃棄低減への応用.....	41
4.1 観測打ち切りへの対応.....	42
4.2 新聞売り子問題とその拡張.....	44
4.3 各店舗各商品の販売時系列での効果検証.....	50
4.4 多店舗多商品の集計販売時系列での効果検証.....	66
4.5 まとめ.....	71
第5章 販売データの統計的性質とその起源の解明.....	72
5.1 べき分布の従来研究.....	73
5.2 商品の入れ替わりの激しさ.....	74
5.3 定常的な対数一様分布の存在.....	75

5.4 対数一様分布の起源.....	77
5.5 まとめ.....	88
第6章 結言.....	89
6.1 まとめ.....	89
6.2 今後の展望	90

第1章 緒言

本章では、始めに社会、経済のビッグデータとその分析により、従来知りえなかった個々の振る舞いや、その集団としての法則性の理解が可能となっており、社会でデータ利活用が進展していることを概観する。次に、経済物理学ではそのビッグデータを物理学の視点から分析して法則性を見出し、数理モデリングを通じて現象を理解し、知見を社会応用してきたことを述べる。金融市場、企業間取引の研究事例に続いて、本研究の題材である販売データにおける先行研究を概観した上で、本研究の目的を明らかにする。

1.1 背景

社会経済ビッグデータとその利活用の進展

近年の情報化の進展は著しく、パソコンやサーバを始め、スマートフォン、ウェアラブル機器などの様々な電子デバイスが、人々の生活や企業活動に深く根付いた。その結果、個人の行動や、社会、経済の活動に関する膨大な情報がビッグデータとしてありのままに記録されるようになった。そのビッグデータの詳細分析により、従来知りえなかった個々の振る舞いや、その集団としての法則性の理解が可能となり、社会でのデータ利活用が進展した。身近な例として、私たちのインターネット上での行動履歴、例えば検索や WEB サイト閲覧、また商品購入の情報は、各個人のインターネット上の行動ログとしてサーバに記録されている。現実世界での行動履歴もまた、スマートフォン等の位置情報を通じて記録されている。それらのデータの分析により、各個人の嗜好性や行動パターンが理解されている。従来のマーケティングにおいては、大衆全体に向けたコマーシャルなどのマスマーケティングが主流であったが、個人の理解が可能となったことで、顧客一人一人の嗜好や行動に合わせてお勧め商品を広告したり、適切な文面とタイミングでダイレクトメールを送付したりする、いわゆる One to One マーケティング [1]が普及した。

各個人のデータを集団としてマクロに俯瞰した時に見えてくる法則性もデータから理解されている。例えば、金融業においては大規模な顧客データを分析して、生命保険の解約やローンの借り換え、住宅の購入などが発生する顧客属性のパターンを見出す取り組みがな

されている [2]。クレジットカードの不正使用パターンの検知や、リスク細分型の自動車保険の顧客属性パターンの抽出、介護保険制度の要介護認定の一次判定などでも同様な取り組みが行われている。

静的な法則性の理解だけでなく、時々刻々と変化する現象を時系列データから理解し、適切な介入によって不都合な事態を回避しようとする取り組みもなされている。医療、健康の分野においては、呼気ガスセンサーなどのウェアラブルデバイスのデータから癌などの病気の予兆を検知しようとする研究がある [3]。振動、音響などの各種センサーを様々な対象に取り付けてネットワーク経由でデータ取得する Internet of Things (IoT) は広く活用検討されており、その時系列データ分析により、牛などの動物の健康状況の把握や [4]、工場の製造設備 [5]、水車発電機 [6]の劣化予兆を捉える研究などがなされている。

経済物理学

社会、経済の膨大なデータは、学術的な研究の対象ともなっている。1990年代に誕生した経済物理学は、経済現象を物理学と同様の視点で実証的に分析し、経済データの中に眠るマクロな統計的性質や法則性を明らかにし、その現象の背後にあるミクロな過程を数理モデリングを通じて理解する。さらに得られた知見を実社会に応用することを特徴とする。具体的な研究事例をいくつか述べる。

金融市場データは経済物理学の代表的な研究題材である。為替取引データベースには市場が開かれる24時間、ミリ秒単位での売り、買い、キャンセルの全ての注文情報が記録されている。その高頻度取引データの分析により、金融市場の詳細な振る舞いが明らかにされてきた。取引価格の変動は従来ランダムウォークとされてきたが、短い時間スケールで大きな価格変化が出現する確率はランダムウォークで想定する正規分布より裾野の広いべき分布に従うという統計的性質が明らかにされた [7]。最新の価格変化に対する過去の価格変化の影響の存在、つまりランダムウォークでは説明できない自己相関があることも確認された [8]。このような様々な経験則がデータから明らかにされ、それらの経験則を再現する時系列モデルとして PUCK (Potentials of Unbalanced Complex Kinetics) モデルが Takayasu らによって提案された [9, 10]。PUCK モデルは市場の不安定さをポテンシャルとして定量化するため、暴騰暴落といった市場不安定性の早期検知や、介入の効果の考察などに活用できると見込まれている [11]。PUCK モデルは市場取引の実務家向けのブルームバーク端末に実装されており、実社会において市場状態分析に活用されている [12]。

同様な事例は、企業間取引データの分析においても見ることが出来る。約 100 万社の企業間取引関係のネットワーク分析により、統計的性質として、各企業の取引数はべき分布しており、取引数は指数関数的に増えることが明らかにされた [13]。そのミクロな起源は、企業の新規参入、倒産、合併がそれぞれ一定の確率で起こることにあり、物質科学におけるエアロゾルの凝集のアナロジーで理解されるものと解明された [13]。また、2 つの企業間の取引金額はそれら企業の売上の積になるという重力則が成立しており [14]、ミクロな相互作用である重力則とネットワーク構造の情報から、企業の年間売上や企業間の取引金額を概ね再現する非線形輸送モデルが構築された [15]。このモデルは、東日本大震災においては、企業売上に対する被害見積もり [16] に活用されるとともに、内閣府の地域経済分析システム RESAS に実装されており、実社会における、持続可能な産業の発展シナリオの考察などへの応用が期待されている [17]。

本研究で題材とする販売データもまた経済物理学の研究の対象となっている。以下、本研究と関連が深い、販売データの統計的性質、時系列分析とその応用について従来研究を概観する。

販売データの統計的性質の従来研究

販売データにおいては、様々な業態や商品について売上、販売数などの統計的性質が調べられている。例えば、スーパーマーケットの調味料の販売数の研究からは、販売数の変化の分布はレヴィ分布で近似でき、裾野がべき分布に従って減衰すること、また自己相関が長期に渡ることが確認されており、金融市場の価格変化との類似性が指摘されている [18]。その統計的性質の起源として広告やソーシャルネットワークの効果が議論されている [19]。Amazon の本の売上 [20] を調べた研究からは、売上の推移がピークに向かう時、またピークから離れる時には、その売上の推移はべき分布に従うということが明らかにされている。コンビニエンスストアのレシート毎の販売額の分布 [21] もべき分布に従うと報告されている。スーパーマーケットの販売データにおいて、価格と販売数との関係がべき関数で記述可能であったという報告もある [22]。このような売上や販売数に関するべき分布の報告がある一方で、デジタルカメラの販売数 [23] は、べき分布に加えて対数正規分布に従う場合があると報告されている。オンラインマーケット [24] においては、価格変化はべき分布でなく指数分布に従っており、価格変化に自己相関がみられなかったという報告もある。

その他、コンビニエンスストアの顧客毎の購買分析からは、25%の顧客の売上が売上全

体の80%を占めていることが確認され、20:80のパレートの法則が概ね成立すると報告されている [25]。また、販売価格については対数正規分布に従っていることが確認されており [26]、商品の販売寿命については、指数分布に加えワイブル分布に従う場合があると報告されている [27]。

販売データの揺らぎの統計的性質を調べた研究からは、販売は一般にポアソン過程でありながら、コンビニエンスストアの商品の販売数の平均と標準偏差の間には、平均が大きい時には標準偏差が平均に比例するというテイラースケーリング則が成立していることが確認されており、その起源は客数の変動にあることが明らかにされている [28]。

これらの統計的性質の応用について、販売額がべき分布することを利用して、高額の支払者の人数と金額から、少額の支払者の人数とその日の売上を推定する提案 [21]や、商品寿命が指数分布でなくワイブル分布に従う場合には、過去の販売日数と残りの商品寿命に相関があることを利用して、残りの販売寿命を予測する提案がある [27]。

販売時系列分析の従来研究

販売時系列の分析については、実務応用の観点から、需要推定の研究が多くなされている。様々な販売時系列、例えば、牛乳 [29]、パン [30, 31]、コンビニエンスストアの食品 [32, 33]、生鮮食品店 [34]の販売時系列を使った需要推定の研究がある。

販売時系列からの需要推定は、販売が基本的にポアソン過程であることから、ポアソン時系列におけるポアソンパラメータ、つまり平均値 λ の推定の問題であると見做せる。さらに、商品の需要は、政治経済、流行、環境変化といった様々な因子の複雑な相互作用により予測不能な変化をすることから、需要推定は非定常なポアソン時系列から平均値 λ を推定するという問題であるといえる。

非定常ポアソン時系列から平均値 λ を推定する従来手法としては、古典的には指数加重移動平均 (Exponentially Weighted Moving Average: EWMA) [35]などの移動平均を基本とした時系列分析手法があり、こういった手法は平均値の推移を大まかに知るために用いることができる。自己回帰和分移動平均モデル (Auto Regressive Integrated Moving Average : ARIMA) などの Box-Jenkins 法のモデル [36, 37, 38, 39]は、自己相関や潜在的な規則性がある非定常時系列の分析を可能とする。一方で、規則性がない、あるいは過去の傾向に従った変化をすることは限らない非定常時系列には適用することができない。このことは、潜在的な規則性を前提とするニューラルネットワーク (Neural Networks : NN)

を始めとする機械学習の手法 [40, 41, 42]でも同様である。一般化線形モデルのポアソン回帰 [43]も、平均値 λ が線形な関数に従って変化することを前提とする。平滑化スプライン (Smoothing Splines: SS) 及び一般化加法モデル (Generalized Additive Mode : GAM) [44]は、変化の規則性を仮定せず、非定常時系列に当てはまる平滑化曲線を求めることで平均値推移の傾向性を明らかにする。一方、平均値 λ の推移が滑らかでない非定常時系列の分析には適さない。状態空間モデル (State Space Model : SSM) [45]は、パラメータの推移がその一期前の値にのみ依存するというマルコフ性を仮定し、パラメータ推定を観測値の尤度を基準に行うため、規則性や滑らかな変化を仮定せずに、非定常時系列のパラメータ推定を実現できる。そのため、非定常ポアソン時系列における平均値 λ 推定の基本モデルとして有用である。

需要推定の問題と関連して、需要に影響する様々な要因の効果を調べる研究も幅広く行われている。例えば、売上に対する天気、曜日、イベント等の各要因の効果を分析した研究 [46]がある。販売促進活動の売上に対する効果について、特別陳列やチラシの効果の時間的な変化や [47, 48]、複数の広告手段の効果の推移を評価した事例がある [49]。商品の販売価格と販売数の関係が、新製品の市場投入によってどのように変化したかを調査した研究もある [50]。顧客の商品の嗜好や選択が変化していく様子を調べた研究 [51]もある。大量陳列などの値引き以外の販売促進活動の有無を、販売データから推定することを試みた研究もある [52]。その他、ある国に対する世論が、その国からの輸入品の需要に与える影響をソーシャルメディアと販売データの関係から評価した事例 [53]がある。販売データから顧客をセグメント化、また商品をカテゴリ化して顧客の購買行動を予測し、需要推定や顧客満足度向上に活用する研究もある [54]。

需要からの仕入決定の従来研究

需要推定に加えて、販売時系列から最大利益を得るための仕入数を決定する研究がある。一般に、小売店等にとって、生鮮食品や新聞など、販売できる期間が短い商品の仕入決定は難しい課題である。商品を仕入れ過ぎて、販売できる期間を超えて商品が売れ残ると廃棄損失となる。一方、仕入れが足りずに商品が売り切れると機会損失となる。需要は一定でなく、ある確率分布に従って揺らぐ場合に、どのような仕入数が利益を最大とするかは自明でない。需要の確率分布を考慮しながら、機会損失と廃棄損失を最適化して、最大利益が得られる仕入数を決める新聞売り子問題は広く研究されている [55, 56, 57]。需要の確率分布を

販売データが得られる度にベイズ更新 [58]の考え方で更新しながら、新聞売り子問題の公式を適用して仕入数を求めるベイジアン動的計画法 (Bayesian Dynamic Programming: BDP) は 1960 年頃から研究されている [59, 60, 61]。

商品の売り切れによって販売数の観測が仕入数で打ち切られた時には、その影響を考慮しなければ需要が過小評価されること [62, 63]が指摘された後は、打ち切りデータに対応した BDP [64, 65]が提案された。打ち切りデータに対応した BDP は、需要が指数分布かワイブル分布の時のみ解けるという制約があったため、近似的な解を求める myoptic 法 [66, 67]も提案された。この myoptic 法は、最適仕入数より大きい仕入数を解として返すことが指摘された [68, 69]ものの、その誤差は多くの場合 1%以下と小さいことが確認された [70]。需要分布の仮定を必要としないノンパラメトリックな手法 [71, 72, 73]も提案されている。このように数々の研究がなされてきたものの、これらの手法が適用できるのは、定常的な需要の場合に限られている。時系列の中に一つだけ変化点を仮定した場合の BDP の研究事例 [74]はあるものの、この手法は打ち切られた販売データからの需要推定には対応していない。非定常で打ち切りがある販売データからの需要推定と仕入決定は、残された研究課題となっている [75]。

小売への社会的な要請

上述の販売時系列分析による需要推定と仕入決定の研究は、最大利益の追求が目的となっている。一方、近年は社会と企業の持続的な発展の観点から、財務情報に現れない環境、社会、企業統治に対する取り組みが企業の社会的責任として求められ、投資家の企業評価の指標にもなっている。環境に関しては、気候変動や資源枯渇などの問題に対して商品や原材料、エネルギーを無駄なく利用すること、社会については、世界的な貧困、人権問題、少子高齢化などの問題に配慮した経営をすること、企業統治については透明性の高い経営と、法令や社会規範を遵守する仕組みを強化することが求められている。

商品、原材料、エネルギーのムダのない利用に関して、本研究と関連が深い食品廃棄低減について詳しく述べる。食品廃棄低減が求められる背景には、世界人口の急速な増加が進んでいることがある。国際連合は世界人口は 2030 年に 86 億人、2050 年に 98 億人に達すると予測している [76]。それに伴い食糧需要が増加するため、食の確保に向けた取り組みが急がれている。その一方で、食品の 1/3 が毎年失われたり廃棄されたりしている現状があり、これは世界中の飢えた人に提供するのに必要な食料の 4 倍に相当すると指摘されて

いる [77]。食品ロスは主に発展途上国における不十分な収穫、貯蔵施設、また食品市場への収穫物の供給網の不備のために発生している [78]。一方、食品廃棄は主に先進国の食品関連産業や家庭での食品廃棄を許容する慣習により発生しているとされている [79]。食品廃棄の低減は、持続可能な開発目標（Sustainable Development Goals: SDGs）のターゲット 12.3 [80]において、「2030年までに小売・消費レベルにおける世界全体の一人当たりの食品廃棄物を半減させ、収穫後損失などの生産・サプライチェーンにおける食品の損失を減少させること」と目標が掲げられている。小売にも具体的な食品廃棄低減に向けた取り組みが求められている。

食品廃棄低減の従来研究

食品廃棄低減に関する従来研究や小売での取り組みを概観する。食品廃棄の構成要素分析 [81, 82, 83]によると食品廃棄の発生源は家庭が最も多く 53-71%、続いて食品製造業の 17-30%、食品関連サービス業 9-12%、そして小売店 2-9%である。小売店の食品廃棄が全体に占める割合は小さいが、SDGs のターゲット 12.3 で示されているように、小売店を含む包括的な食品廃棄低減の取り組みが求められている。小売店での食品廃棄の構成要素分析からは、パン、フルーツ、野菜といった傷みやすい食品が廃棄の主要な要素であるとされている [84, 85]。調査事例の一つによると小売店の食品廃棄は、重量換算で 30%がパン、29%がフルーツと野菜によるものとされている [86]。

食品廃棄低減に向けて多くの方法が検討されている。廃棄される可能性がある食品の再配布、寄付の取り組みは先進国で広く普及している [87, 88]。近年、ベンチャー企業がオンラインプラットフォームにて、余剰食品の提供 [89, 90]や、調理や食事を共にする [91, 92]ためのサービスを提供している。他の主要な手法として、廃棄食品を化学肥料やエネルギーに変えたり、廃棄食品から特定の物質を抽出したりする食品廃棄の転換 [93, 94]がある。熱化学、バイオ科学、発酵、抽出などの手法が、廃棄食品から燃料、薬、化粧品、家畜の餌、肥料を得ることに使われている。その他、様々な啓蒙キャンペーン、Love Food HateWaste [95]や I Value Food [96]などが、顧客の自発的な食品廃棄低減に向けて実施されている。

実際に小売店で実施されている廃棄低減の取り組みも調査されている。イギリスの生鮮食品店では、慈善事業やフードバンクへの食品寄付が、余剰食品の供給プラットフォームである FareShare [97]を活用して実施されている [98]。オンラインプラットフォーム Food

Cloud [99]は期限切れの商品を把握したり、登録された慈善事業団体との情報交換を容易にしたりして、食品寄付を簡易化することに貢献している。その他、賞味期限の表示を廃止して商品をパックした日だけ表示したり [100, 101]、食品品質を長く保つためにパック方法や貯蔵庫の改良をしたり、販売期限が近づいた商品や販売基準に満たない形状の商品を値引き販売したりする取り組みがなされている。他の事例として、余剰食品を店員に提供すること、パンなどの廃棄物を家畜の餌にすること、調理用油をバイオ燃料の生成に活かすことも報告されている。

廃棄低減の施策導入は、経営的また技術的な観点を十分考慮しなければ失敗することが指摘されている [102]。食品廃棄低減の施策と、経営者からの評価が調査されており [103]、スイスのレストランでは、既存の業務プロセスに既存の技術を使って行う廃棄低減の取り組みが、新規性の高い技術を使う取り組みより導入が進んでいた。新技術の導入の制約は幅広く存在しており、内的要因（インフラ、スペース等）、人的要因（店員の数、スキル、モチベーション等）、外的要因（顧客や供給者からの協力等）、法的要因（法律や条例等）などが挙げられている。経営者の新技術導入の動機は、経済的な恩恵とブランドイメージ向上が主要因となっている。こういった観点を考慮して、実用的な廃棄低減の提案をする必要がある。

なお、販売データを使った廃棄低減の研究としては、廃棄の量を販売データを使って定量化し、リフィルパッケージを導入することによるシャンプーやヘアコンディショナーの容器の廃棄低減量を求めた事例がある [104]。その他の販売データ研究は一般に上述のとおり利益向上を主としており、廃棄低減に注目した研究は殆ど見られない。

1.2 目的

販売データの統計的性質、時系列分析とその応用の従来研究、また小売への廃棄低減の社会的要請を鑑みると、いくつかの重要課題があることが分かる。

販売データの統計的性質として、平均が大きい時に平均と標準偏差の関係がポアソン過程から乖離するというテイラースケール則の知見が得られていながら、需要推定を始めとする非定常ポアソン時系列のパラメータ推定において、その考慮がなされていない。非定常時系列においてデータと整合した正しいパラメータ推定を行うためには、観測データの変動があった時、それをパラメータの非定常な変動と見做すのか、観測値の揺らぎとみな

すのか、テイラースケーリングを考慮しながら正しく区別する必要がある。本研究ではテイラースケーリング則を考慮した非定常ポアソン時系列のパラメータ推定手法を確立することを一つ目の目的とする。この手法の確立により、販売時系列からの精緻な需要推定を実現できると期待される。そのことに加え、テイラースケーリング則は、生態学 [105]、金融市場 [106]、ネットワーク [107]、インターネットのブログでの単語の出現数 [108]など、様々な自然、社会の現象で確認されていることから、この手法の確立は、テイラースケーリング則に従う様々な現象の非定常ポアソン時系列のパラメータ推定に寄与すると見込まれる。

このような販売時系列からの需要推定を応用し、適切な仕入を実現することは、小売店での廃棄低減の課題に対して効果的、かつ小売店で実施可能な施策と考えられるものでありながら、その具体的な提案は見られない。上述のとおり、非定常で売り切れを含む販売時系列から需要推定と仕入決定を行う手法が存在しないことが一つの課題である。併せて、仕入決定に際しては従来のような最大利益の追求だけでなく、利益減少を最小限にとどめながら廃棄低減できる仕入決定手法の確立が求められる。本研究では上述の非定常ポアソン時系列のパラメータ推定手法を拡張してこれらの課題に対応させ、消費期限が短い商品の廃棄低減を、利益減少を最小限に留めながら実現する販売データ分析手法を提案することを二つ目の目的とする。

さらに販売データの統計的性質に関して、上述のテイラースケーリング則や、売上のべき分布などいくつかの性質が明らかにされている一方で、企業の機密情報にあたる詳細な販売データが学術的に研究される事例は多くなく、未だ明らかとなっていない販売データの統計的性質が存在する可能性がある。本研究では、コンビニエンスストア 300 店舗あまりの5か月間に渡るレジでの全販売記録 1.6 億レコードを活用することにより、販売データの統計的性質を調査する。その調査の結果から、商品の販売店舗数が対数一様分布に従うという新たな統計的性質を報告するとともに、その起源を解明することを三つ目の目的とする。

本論文の構成

上述の目的をまとめた本論文の構成を図 1.1 に示す。本論文では、1 店舗 1 商品から多店舗多商品まで、様々なスケールでの販売データの数理科学の発展に寄与する。販売データの統計的性質といった基礎研究に加え、廃棄低減などの現実の問題への応用研究も実施する。

本章に続く第 2 章では、本研究で用いるコンビニエンスストアの販売データについて説明する。第 3 章では、テイラースケーリング則によるポアソン過程からの標準偏差の乖離を考慮した 1 店舗 1 商品の販売時系列の非定常パラメータ推定手法を確立する。第 4 章では、第 3 章の時系列分析手法を拡張して、小売店での需要推定と仕入決定に適用可能とし、利益減少を最小限にとどめながら廃棄低減する手法を確立する。1 店舗 1 商品の販売時系列から、多店舗多商品の集計した販売時系列の場合までの廃棄低減の効果検証を行う。第 5 章では、多店舗多商品の販売データ分析から見出した新たな統計的性質として、商品の販売店舗数の分布が対数一様分布に従うことを示す。その数理的な起源を、1 商品の販売店舗数が従う確率過程の考察から明らかにする。第 6 章で本論文を総括し今後の展望を述べる。

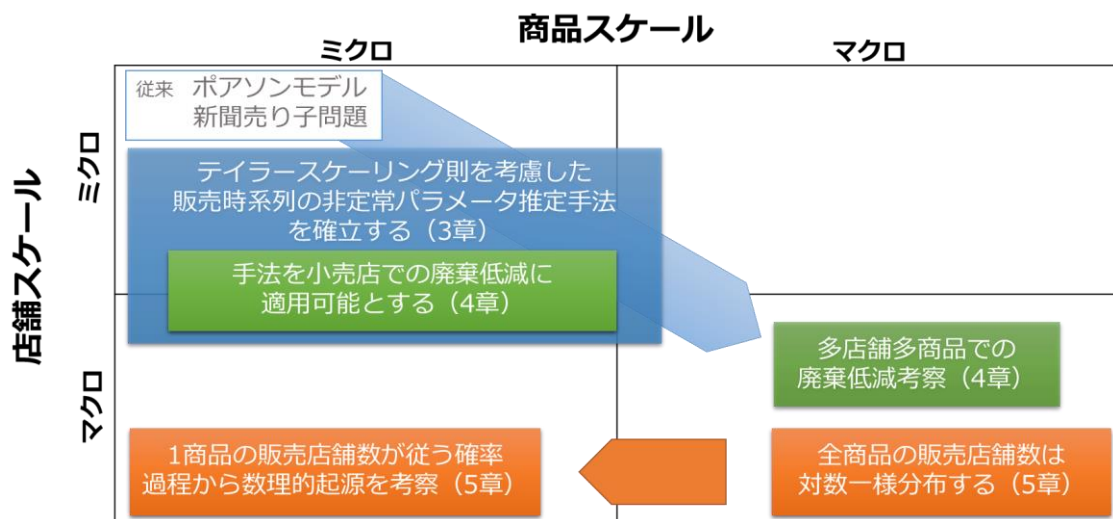


図 1.1 本論文の構成

第2章 データ

本章では、始めに販売時点情報（Point-Of-Sales : POS）の一般的な概要を述べる。次に本研究で用いるコンビニエンスストアの POS データについて説明する。326 店舗の 5 か月間に渡る約 1.6 億レコードのレジでの販売記録に加え、日々の仕入と廃棄の記録が店舗毎、商品毎にあることを中心に述べる。

2.1 販売時点情報（Point-Of-Sales : POS）の概要

POS データは販売時点、つまり店舗のレジで支払いがなされる時に記録される情報である。商品値札のバーコード（日本では JAN コード）をスキャナーで読み取ることで、いつ、どの店舗で、どの商品が、いくらで、何個売れたのかが記録される。POS データは売上の集計作業を容易化し、販売戦略の立案に欠かせない売上分析の効率化、高精度化を実現する。店舗での一般的な POS データ活用方法としては、売れ筋、死に筋の商品の特定があり、売れ行きを ABC ランク分けしてランク毎に仕入調整するための ABC 分析や、売れる時期を見極めて仕入調整するためのトレンド分析がある。顧客を商品購入の時期、頻度、金額を基準にセグメントに分けて販売施策を検討する RFM 分析も行われる [109]。ポイントカードなどの顧客 ID が記録された POS は ID-POS と呼ばれる。その分析により、どんな年代、性別の人がどのような商品を購入しているか、商品のリピート率なども含めて分析される。このように POS データは、顧客がどのような商品をどのくらい求めているのかを理解することを助け、売上向上に役立てられている。

POS のシステムは、日本では 1982 年にセブンイレブンが POS レジシステムを初めて導入して以来、コンビニエンスストア、スーパーマーケット、ドラッグストアなどの小売店に幅広く導入されてきた。近年では大規模なチェーン店のみならず、個人商店でも POS を導入しやすくするための様々なソリューションが提案されている [110]。従来の大規模チェーンストア向けのレジ一体型のターミナル POS は、1 台当たり初期費用 50 万円、月額 1 万円といった費用が必要であったのに対し、PC ベースの POS は比較的安価で 1 台当たり初期費用 20 万円、月額 5 千円程度、タブレットなどのスマートデバイス型 POS は初期費

用、月額ともに0円から提供されるものも登場しており、POSの導入の敷居が下がっている。

POSデータは人々の購買行動や店舗の販売活動を大規模かつ詳細に記録している。このデータを分析することで、人々の行動や販売に関するマクロな統計的性質を理解したり、個人の購買行動や各店舗の販売活動などを明らかに出来ると期待される。

2.2 本研究で用いたPOSデータ

本研究で用いたコンビニエンスストアのPOSデータは、株式会社セブン-イレブン・ジャパンの326店舗で2010年6月1日から10月31日までの153日間で取得されたものである。326店舗の所在地は164店が神奈川県、162店が山口県である。このPOSデータには、各店舗のレジでの販売記録が合計で約1.6億(159,477,360)レコードある。各レコードには、店番、販売日時、商品コード、販売数量、販売金額、商品分類などが記録されており、その一部の約2千万(22,261,929)レコードには、電子マネーEdyのIDが約30万(324,188)個分、記録されている。販売と併せて、日々の仕入と廃棄の情報が店舗毎、商品毎にある。また、販売促進活動の情報として、販売促進対象の商品コード、販売促進の開始終了日、値引き額や値引き率の情報がある。その他、JANコードマスタがあり、商品コード毎の商品名、JANコード、推奨売価などの情報がある。なお、商品コード毎の原価の情報は得られていない。

各店舗で取り扱われる商品は、食品では、おにぎり、パン、弁当、冷凍食品、お菓子、ドリンクなどに加え、店舗で調理するフライドチキン、おでんなどがある。食品以外にも、新聞、雑誌、たばこ、文具、生活用品、ペットフードに至る幅広い商品が扱われている。POSデータに記録された商品コードの全数は約2万(21,037)個であり、1店舗あたりの平均の商品コード数は約1600個(最小で約900個、最大で約2200個)である。2010年の店舗の平均売り上げは約4200万円、また食品の売り上げが全売り上げに占める割合は67.4%、店舗の平均敷地面積は123m²である [111]。

商品の廃棄について、廃棄が1度でもあった商品コードの数は6,502個、廃棄総数は約6百万(6,240,363)個であった。この内、消費期限が短い食品である、おにぎり、揚げ物、肉まん類、パン、総菜、麺類、弁当、おでん、パック食品、パック飲料、デザート類に限定すると商品コード数は2,507個、廃棄総数は約600万(6,183,964)個であっ

た。それらの商品の販売総数は約8千万（80,922,819）個であったことから、単純計算で仕入数の7%程度が廃棄されていた計算になる。

表 2.1 本研究で用いた POS データ

項目	内容
店舗数	326店（神奈川県：164店、山口県：162店）
期間	2010/6/1～10/31（153日）
販売記録	約1.6億枚のレシート情報 - 店舗コード、購入日時、商品コード、商品名、単価、販売数 - 約2200万レシートは電子マネーEdyのID（約30万ID）付き
仕入記録	店舗毎、商品毎の日々の仕入記録
廃棄記録	店舗毎、商品毎の日々の廃棄記録
販売促進の記録	販売促進毎の商品コード、開始日、終了日、値引き額/率
JANコードマスタ	商品コード毎の商品名、JANコード、推奨売価等 （原価の情報はなし）

第3章 販売データの統計的性質を考慮した 非定常時系列分析手法の確立

第3章では、販売データの統計的性質を考慮した非定常ポアソン時系列のパラメータ推定手法を確立する。始めに、ポアソン過程の標準偏差が平均値に比例して大きくなるというテイラースケーリング則の概要を述べた上で、非定常な需要を正しく推定するためには、ポアソンパラメータの非定常な変化とテイラースケーリングによる大きな揺らぎとの区別が欠かせないことを指摘する。その対応として、非定常時系列分析手法である粒子フィルタにおいて、観測モデルにテイラースケーリング則を考慮した尤度関数を導入する。さらに現実の販売時系列においては地域のイベント等による突発的な需要の急増が発生するため、非連続的な需要の変動を推定する必要があることに着目し、非連続性の検知と推定値の補正の方法を開発し粒子フィルタに導入する。提案手法の効果を、乱数で人工的に生成した時系列とPOSの販売時系列で検証し、テイラースケーリング則による標準偏差のポアソン過程からの乖離が存在し、非連続的な変化を含む非定常な販売時系列において、提案手法によってデータと整合する適切な需要の推定が可能となることを確認する。なお、本章の内容は論文 [112]に基づく。

3.1 テイラースケーリング則

テイラーは昆虫の個体数の観察を通じて、平均 λ と標準偏差 σ の間に次式のようなスケーリング則を見出した [113]。

$$\sigma = \gamma \cdot \lambda^\beta \quad (3.1)$$

ここで、 γ は比例定数、 β はべき指数であり、イベントが無相関の場合は $\beta = 1$ となつて、標準偏差は平均に比例する。イベントに相関がある場合には β が0.5から1となることが知られている。その具体例として販売データにおいては、発売日に多く買われる雑誌や新聞、まとめ買いされる切手で、1以下の β が確認されている [28]。

テイラースケーリング則は表 3.1 に示したように自然、社会の様々な現象で確認されている [114]。宇宙線の照射量、生物の個体数、細胞の個数、鳥の群れの大きさなど、単位面積当たりで考える事象 (Ensemble Fluctuation Scaling)、植物の成長、河川の流量、高速道路の通行量、株の取引高、商品販売数の時系列のような単位時間当たりで考える事象 (Temporal Fluctuation Scaling) の両方において、テイラースケーリング則の成立が確認されている。

テイラースケーリング則を再現するモデルとして、ランダム拡散モデル (Random Diffusion Model : RD モデル) が知られている [107]。RD モデルでは、多数のノードがそれぞれ様々なリンク数で繋がる複雑ネットワークを想定し、そのネットワーク上をランダムウォーカーが移動する過程を考える。ランダムウォーカーは毎時間いずれかのノードに追加されるとともに、設定された寿命になると消失するものとする。各ノードの単位時間当たりのランダムウォーカーの通過数の時系列で、平均と標準偏差の関係を調べると、テイラースケーリング則が再現される。RD モデルは解析的にも解かれており、テイラースケーリングの起源はネットワーク全体に存在するランダムウォーカーの母数が、追加と消失によって揺らぐことにあると解明されている [115]。コンビニエンスストアの販売データにおけるテイラースケーリングの起源は、客数の揺らぎにあると示されている [28]。

イベントに相関がなく $\beta = 1$ の時、テイラースケーリングに従うポアソン過程の標準偏差 σ は、平均 λ が小さい時には λ の平方根となり、 λ が大きい時には λ に比例することから、次の式で記述される。

$$\sigma = \sqrt{\lambda + (\gamma \cdot \lambda)^2} \quad (3.2)$$

表 3.1 テイラースケーリング則の成立が確認されている現象（文献 [114]をもとに再構成）

Subject	System
Networks	Random walk , Network models , Highway network , World Wide Web , Internet
Physics	Heavy-ion collisions, Cosmic rays
Social	Stock market
sciences/economics	Stock market, Business firm growth rates, Email traffic, Printing activity
Climatology	River flow , Precipitation
Ecology/ population dynamics	Forest reproductive rates , Satake–Iwasa forest model , Crop yield , Animal populations , Diffusion Limited population, Population growth , Exponential dispersion models, Interacting population model
Life sciences	Cell numbers, Protein expression , Gene expression , Individual health, Tumor cells, Human genome, Blood flow, Oncology, Epidemiology

テイラースケーリングの比例定数 γ の決定

式3.2中の比例定数 γ の値は、実データの平均と標準偏差の関係を調べることで決定できる。図 3.1 は、あるソフトドリンクの平均と標準偏差の関係を調べたものである。黒一点鎖線は $\sigma = \sqrt{\lambda}$ 、黒点線は $\sigma = \gamma \cdot \lambda$ 、黒実線は $\gamma = 0.12$ とした時の式3.2の関係である。

図 3.1 中の灰色点は次の手順により求めた。

- 平均が小さい側の平均と標準偏差の値を多く得るため、対象のソフトドリンクの各店舗の販売データをそれぞれ曜日毎に分けて、各々の平均と標準偏差を求める。この際、販売記録に欠損があるデータは除外する。
- 平均が大きい側の平均と標準偏差の値を得るため、上記の曜日毎の各店舗の販売データを曜日毎にランダムにサンプリングし、日々の販売数を合算した上で平均と標準偏差を計算する。ここで曜日毎のランダムサンプリングは、サンプリングする販売データの数を 2 から販売店舗数上限まで 1 ずつ上昇させながら、各 30 セットずつ乱数シードを変えて実施する。

以上の手順により、図 3.1 中の灰色点 49924 点を得た。比例定数 γ の値はこのデータを式3.2で非線形回帰 [116]して 0.12 と決定できる。統計分析ソフトウェア R の nls 関数 [117]を用いることで、式3.2を含む任意の関数形による非線形回帰を容易に実行できる。

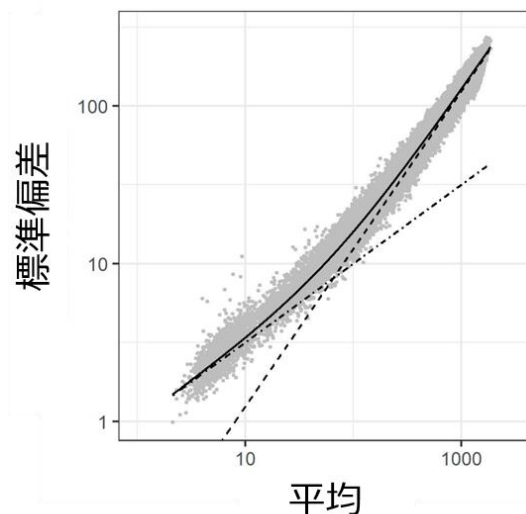


図 3.1 あるソフトドリンクの販売データにみられるテイラースケーリング。黒線は $\sigma = \sqrt{\lambda + (\gamma \cdot \lambda)^2}$ 、黒一点鎖線は $\sigma = \sqrt{\lambda}$ 、黒点線は $\sigma = \gamma \cdot \lambda$ 。 $\gamma = 0.12$ 。

従来のパラメータ推定手法

テイラースケーリング則に従うデータの従来のパラメータ推定手法について記された文献がある [118]。この文献では、植物の成長期において、植物の重さの推移を調査する場合が例示されている。ここでは、 q 回の観測を行い、各観測で n_i 個、合計 $n = \sum_{i=1}^q n_i$ 個のデータを得るとし、線形または非線形の成長関数 $f(x; \theta)$ を見積もるものとしている。 x は説明変数、 θ は推定対象のパラメータのベクトルである。例えば、 θ のパラメータ数 $p = 2$ で、 $f(x; \theta)$ が線形ならば $f(x; (a, b)) = a + bx$ である。 $\mu_i = f(x_i; \theta)$ を予測値、 y_{ij} はグループ i での j 個目の観測値、 $\hat{\theta}$ をパラメータ推定結果として、次の5つがパラメータ推定に使われると記されている。

a) 最小二乗法

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^q \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 \quad (3.3)$$

b) χ^2 最小化

$$\hat{\theta} = \operatorname{argmin}_{\theta} \chi^2 = \operatorname{argmin}_{\theta} \sum_{i=1}^q \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_i)^2}{|\mu_i|} \quad (3.4)$$

c) ψ^2 最小化

$$\hat{\theta} = \operatorname{argmin}_{\theta} \psi^2 = \operatorname{argmin}_{\theta} \sum_{i=1}^q \left[\frac{\frac{\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{n_i - \frac{n_i}{n} p}}{\left(\frac{\mu_i}{\sqrt[q]{\prod_{k=1}^q \mu_k}} \right)^{\beta}} \right] \quad (3.5)$$

ここで β は、分散 V と平均 μ の関係 $V_i = \alpha \mu_i^{\beta}$ で定義されるテイラースケーリングの指数である。

d) 重み付き最小二乗法

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^q \left[w_i \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 \right] \quad (3.6)$$

$$w_i = \left[\sum_{j=1}^{n_i} \frac{(y_{ij} - \bar{y}_i)^2}{n_i - 1} \right]^{-1} \quad (3.7)$$

ここで、 \bar{y}_i はグループ i での観測値の平均値である。

e) 最尤推定法

$$\begin{aligned} \hat{\theta} &= \operatorname{argmax}_{\theta} \ln(L) \\ &= \operatorname{argmax}_{\theta} \left\{ - \left[\frac{n}{2} \ln(2\pi\alpha) + \frac{\beta}{2} \sum_{i=1}^q n_i \ln(\mu_i) + \sum_{i=1}^q \frac{\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{2\alpha\mu_i^{\beta}} \right] \right\} \end{aligned} \quad (3.8)$$

ここで L は下記の尤度関数であり、この文献 [118]では正規分布が仮定されている。

$$L(\mu_i; \alpha, \beta) = \prod_{i=1}^q \prod_{j=1}^{n_i} \left\{ \frac{1}{\sqrt{2\pi\alpha\mu_i^{\beta}}} \exp \left[- \frac{(y_{ij} - \mu_i)^2}{2\alpha\mu_i^{\beta}} \right] \right\} \quad (3.9)$$

上記 a)から e) のいずれの方法でも関数 $f(x; \theta)$ は、線形または非線形の何らかの関数で規定できることが仮定されている。本研究で対象とする、関数形が想定できない複雑な非定常性がある場合のテイラースケーリング則に従うポアソン時系列のパラメータ推定に対して、上記の従来手法は適用できない。非定常なポアソン時系列のパラメータを正しく推定するためには、観測データの変動があった時、それをパラメータの非定常な変動と見做すのか、観測値の揺らぎとみなすのか、テイラースケーリングを考慮しながら正しく区別する必要がある。一方で、現状はそれを実現できる非定常時系列分析手法の提案がない。

3.2 状態空間モデル

提案する非定常時系列分析手法は状態空間モデルを基礎としている。図 3.2(a)に示したように状態空間モデルでは、観測できない状態 x_t を想定し、時間 t における観測値 y_t は、状態 x_t からある確率分布に従って生じると考える。状態 x_t はひとつ前の時間の状態 x_{t-1} に依存して確率的に決まるとする。

$$x_t \sim \Phi(x_t | x_{t-1}) \quad (3.10)$$

$$y_t \sim \Phi(y_t | x_t) \quad (3.11)$$

$\Phi(A|B)$ は B で条件付けされた A の確率密度関数(Probability Density Function: PDF)であり、チルダ(\sim)は左辺が右辺のPDFの実現値であることを意味する。式3.10は状態 x_t の時間発展を記述するものであり、システムモデルと呼ばれる。式3.11は状態 x_t からの観測値 y_t の確率的な生成を意味しており、観測モデルと呼ばれる。ひとつ前の状態に依存して次の状態を推定するため、データ推移の関数形や、自己相関などの潜在的な規則性の仮定を必要としない。現実の系の時系列で想定される予測できない非定常性や急峻な変化があってもパラメータ推定に適用できる。

現実の時系列と状態空間モデルとの対応を説明するため、簡単な例として図 3.2(b)に乱数で作成したポアソン時系列を示す。赤線は各時点での平均値 λ であり、観測されない状態 x_t に対応する。緑線は、各時点での平均値 λ を仮定して得たポアソン乱数であり観測値 y_t に対応する。ここで、観測値 y_t は各時点での状態 x_t 、つまり平均値 λ に応じたポアソン分布から確率的に発生していると見做している。状態空間モデルでは、このように状態 x_t から観測値 y_t が得られるとしたモデリングを行う。実際の時系列分析では、得られた観測値 y_t から状態 x_t (平均値 λ)を推定する。その方法の基礎となる、状態空間モデルによる状態の逐次推定の方法について以下に述べる。

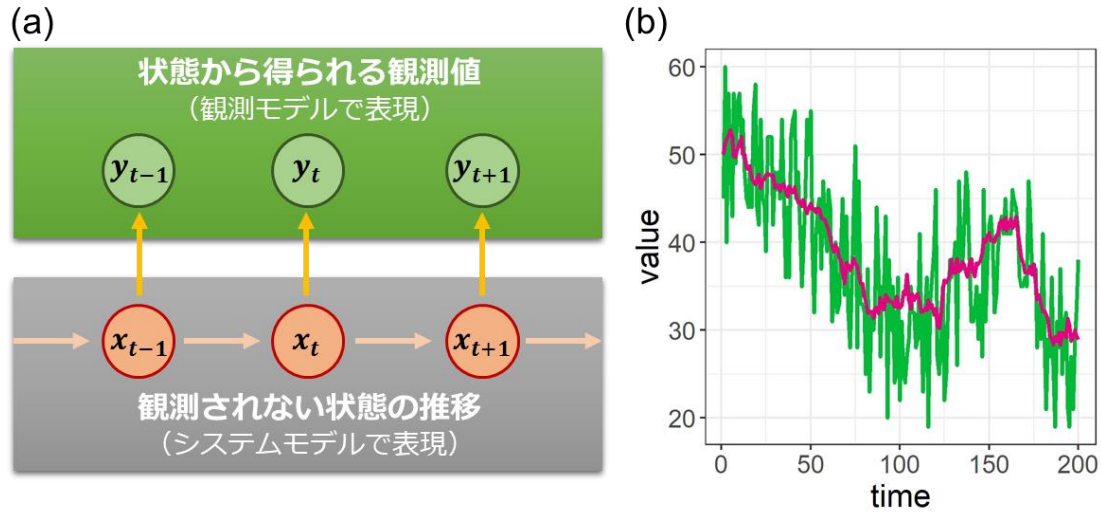


図 3.2 状態空間モデル概要の説明図。(a) 状態から観測値が得られる概念図。(b) 乱数で作成したポアソン時系列。赤線は各時点の平均値 λ であり状態に対応、緑線は各時点での平均値 λ を仮定して得たポアソン乱数で観測値に対応。

状態の逐次推定

$t - 1$ 時点までの観測値 $[y_1, y_2, \dots, y_{t-1}]$ を $y_{1:t-1}$ とする。 $y_{1:t-1}$ が得られている条件下での x_t の予測分布 $p(x_t|y_{1:t-1})$ は次のように書ける [119]。

$$\begin{aligned}
 p(x_t|y_{1:t-1}) &= \int p(x_t, x_{t-1}|y_{1:t-1}) dx_{t-1} \\
 &= \int p(x_t|x_{t-1}, y_{1:t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1} \\
 &= \int p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1} \tag{3.12}
 \end{aligned}$$

$p(x_t|x_{t-1})$ は前述の式3.10のシステムモデルに相当する。 $p(x_{t-1}|y_{1:t-1})$ はフィルタ分布と呼ばれ、 $t - 1$ 時点までの観測値 $y_{1:t-1}$ の尤度から求められる x_{t-1} の確率分布である。なお、式3.12の第1式では周辺化 $p(A) = \int_B p(A, B) dB$ 、第2式への変形では乗法定理 $p(A, B) = p(A|B)p(B)$ 、第3式への変形ではマルコフ性の仮定 $p(x_t|x_{t-1}, y_{1:t-1}) = p(x_t|x_{t-1})$ が用いられている。

t 時点でのフィルタ分布 $p(x_t|y_{1:t})$ は、 $p(x_t|y_{1:t}) = p(x_t|y_t, y_{1:t-1})$ と書けることから、ベイズの定理を用いて、次のように記述できる。

$$\begin{aligned}
 p(x_t|y_{1:t}) &= \frac{p(y_t|x_t, y_{1:t-1})p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \\
 &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \\
 &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t} \tag{3.13}
 \end{aligned}$$

$p(y_t|x_t)$ は式3.11の観測モデルに相当し、 $p(x_t|y_{1:t-1})$ は式3.12の予測分布である。なお、式3.13の第2式への変形ではマルコフ性の仮定 $p(y_t|x_t, y_{1:t-1}) = p(y_t|x_t)$ 、第3式への変形では周辺化が用いられている。

式3.12による一期先の予測と式3.13によるフィルタリングを繰り返すことにより、観測値 y_t が得られた時の状態 x_t の逐次推定を行うことができる。

システムモデル

システムモデルの式3.10は、 $f_t(\cdot)$ を任意の関数、 v_t を任意の確率分布 p に従うノイズとして次のように書ける。

$$x_t = f_t(x_{t-1}, v_t) \tag{3.14}$$

$$v_t \sim p(v|\theta_{sys}) \tag{3.15}$$

ここで、 θ_{sys} は確率分布 p を規定するパラメータである。我々は非定常ポアソン時系列のパラメータ推定のためのシステムモデルを次のように定義した。

$$x_t = \delta(x_{t-1} + v_t) \tag{3.16}$$

$$\delta(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.17}$$

$$v_t \sim (1 - m) \cdot N(0, \alpha \cdot x_{t-1}) + m \cdot U(-\beta, \beta) \tag{3.18}$$

状態 x_t はポアソン分布の平均値 λ に対応するため、その値は正である必要がある。そのため、式3.17のデルタ関数により式3.16の状態 x_t の値を正に制限している。式3.18に示したように、ノイズ v_t は、平均0で標準偏差 $\alpha \cdot x_{t-1}$ の正規分布 $N(0, \alpha \cdot x_{t-1})$ に加えて、 $-\beta$ から β の値をとる一様分布 $U(-\beta, \beta)$ を比率 m で混合した分布に従うとしている。一様分布の β の値を正規分布の揺らぎよりも十分大きくとることで、大きな非定常変化に対応できるようにしている。

このような正規分布と一様分布とを混合したシステムノイズ v_t は Yura らによって提案されている [120]。式3.18中のハイパーパラメータ α, β, m の決定については後述する。

観測モデル

観測モデルの式3.11は、 $h_t(\cdot)$ を任意の関数、 w_t を任意の確率分布 p に従うノイズとして次のように書ける。

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{w}_t) \quad (3.19)$$

$$\mathbf{w}_t \sim p(\mathbf{w} | \theta_{obs}) \quad (3.20)$$

ここで、 θ_{obs} は確率分布 p を規定するパラメータである。従来のポアソンモデルにおける観測モデルは次のように書ける。

$$\mathbf{y}_t \sim \mathbf{Po}(\mathbf{x}_t) \quad (3.21)$$

$\mathbf{Po}(\mathbf{x}_t)$ は平均が \mathbf{x}_t の時のポアソン分布の PDF であり、観測値 \mathbf{y}_t はポアソン分布に従うことを表している。

本研究では、テイラースケーリング則で規定される平均と標準偏差の関係、式3.2を考慮した観測モデルを導入する。ポアソン分布の標準偏差は平均の平方根となるため、テイラースケーリングの平均と標準偏差の関係を考慮できない。そのため、提案する手法では、 \mathbf{x}_t が統計的に大きい場合、例えば \mathbf{x}_t が 20 以上の時にポアソン分布を正規分布で近似し、正規分布の標準偏差にテイラースケーリングの項を組み込むことで、テイラースケーリングを考慮する。具体的には次のような観測モデルを定義する。

$$\mathbf{y}_t \sim \begin{cases} \mathbf{Po}(\mathbf{x}_t), & \text{if } \mathbf{x}_t < 20 \\ N\left(\mathbf{x}_t, \sqrt{\mathbf{x}_t + (\gamma \cdot \mathbf{x}_t)^2}\right), & \text{otherwise} \end{cases} \quad (3.22)$$

ポアソン分布の正規分布近似は $\mathbf{x}_t = 10$ でも実施でき、 $\mathbf{x}_t \geq 20$ の時にはポアソン分布と正規分布は実質的に区別できないことが知られている [121]。 $\mathbf{x}_t = 20$ の時のポアソン分布の歪度は、 $\mathbf{x}_t = 10$ の時より 30%小さく正規分布の歪度 0 に近づく。一方、テイラースケーリングによる標準偏差の変化率は、テイラースケーリングの比例定数 γ が本研究の POS データに典型的な値である 0.1 とした時、 $\mathbf{x}_t < 20$ では 10%未満と小さい。そのため、本研究では $\mathbf{x}_t \geq 20$ の時にポアソン分布を正規分布で近似することとした。

線形・ガウス型状態空間モデルの状態推定

システムモデルの式3.14と観測モデルの式3.19のそれぞれの関数 $f_t(\cdot)$ 、 $h_t(\cdot)$ がともに線形で、かつそれぞれのノイズ v_t 、 w_t がともに正規分布に従う場合には、カルマンフィルタと呼ばれる効率的な状態推定の方法が知られている [122]。この時の状態空間モデルは、

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{G}_t \mathbf{v}_t \quad (3.23)$$

$$\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \quad (3.24)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{w}_t \quad (3.25)$$

$$\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R}_t) \quad (3.26)$$

と書ける。ここで、 \mathbf{Q}_t 、 \mathbf{R}_t は分散共分散行列、 \mathbf{F}_t 、 \mathbf{G}_t 、 \mathbf{H}_t は任意の値をもつ行列である。この時、状態 \mathbf{x}_t は正規分布となり、平均ベクトルと分散共分散行列で規定される。B 時点までの観測データが与えられた時の A 時点での平均ベクトルを $\mathbf{x}_{A|B}$ 、分散共分散行列を $\mathbf{V}_{A|B}$ とすると、カルマンフィルタによる一期先予測の式は、式3.23、式3.24より、

$$\mathbf{x}_{t|t-1} = \mathbf{F}_t \mathbf{x}_{t-1|t-1} \quad (3.27)$$

$$\mathbf{V}_{t|t-1} = \mathbf{F}_t \mathbf{V}_{t-1|t-1} \mathbf{F}_t^T + \mathbf{G}_t \mathbf{Q}_t \mathbf{G}_t^T \quad (3.28)$$

と書ける。観測値 \mathbf{y}_t が与えられた時の t 時点の状態推定は次のように行う。

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_{t|t-1}) \quad (3.29)$$

\mathbf{K}_t はカルマンゲインと呼ばれ、観測値 \mathbf{y}_t と予測値 $\mathbf{H}_t \mathbf{x}_{t|t-1}$ の差を、どの程度状態に反映させるかを規定する。 \mathbf{K}_t は観測モデルの分散共分散行列 \mathbf{R}_t に対し、状態の1期先予測の分散共分散行列 $\mathbf{V}_{t|t-1}$ が相対的にどの程度の大きさかで決められる。

$$\mathbf{K}_t = \mathbf{V}_{t|t-1} \mathbf{H}_t^T (\mathbf{H}_t \mathbf{V}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \quad (3.30)$$

状態の予測誤差を表す分散共分散行列 $\mathbf{V}_{t|t-1}$ が、観測の誤差の分散共分散行列 \mathbf{R}_t に対して小さいならば、 \mathbf{K}_t は小さくなるため、観測値と観測予測値に差は状態にあまり反映されないこととなる。状態の分散共分散行列のフィルタリングは次のように行う。

$$\mathbf{V}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{V}_{t|t-1} \quad (3.31)$$

この式はカルマンゲインが大きく、状態が大きく補正される場合には、状態の予測誤差の分散共分散行列の値を小さくすることを意味する。

非線形・非ガウス型状態空間モデルの状態推定

状態空間モデルが非線形、あるいはノイズ分布が非ガウス型の場合は、状態の分布は正規分布とならない。そのような状態空間モデルでの状態推定方法として、非線形モデルの状態を正規分布で近似する拡張カルマンフィルタや、混合ガウス分布を用いて分布を近似しカルマンフィルタを適用する方法が知られている。一方、拡張カルマンフィルタは分布が多峰性、または非対称な時に良好な結果が得られず、混合ガウス分布を用いる方法は時間の進行とともに正規分布の項数が増え計算量が膨大になる課題がある。状態の分布を階段関数近似する方法も知られているが、数値積分が必要なため、同様に計算量に課題がある [122]。

3.3 粒子フィルタとその拡張

粒子フィルタ

状態の分布を関数で近似するのではなく、任意の値をもつ独立な粒子の密度で近似する方法がある。粒子近似した $p(x_t|y_{1:t})$ をもとにフィルタリングを行う手法は粒子フィルタ [123, 124]と呼ばれる。 $t - 1$ 時点での状態 x_{t-1} が N 個の粒子 $\{f_{t-1}^1, \dots, f_{t-1}^N\}$ の分布で表されているとする。システムノイズ v_t が N 個 $\{v_t^1, \dots, v_t^N\}$ 与えられている時、予測分布は次のように書ける。

$$p_t^j = F_t(f_{t-1}^j, v_t^j) \tag{3.32}$$

観測値 y_t が与えられた時の状態 x_t の推定は次のように行う。最初に N 個の各粒子 p_t^j について、 y_t が得られる尤度 $\beta_t^j = p(y_t|p_t^j)$ を求める。次に各粒子 p_t^j の尤度 β_t^j に比例した確率で各粒子 p_t^j の復元抽出を N 回行う。これにより状態 x_t の分布を表す N 個の粒子 $\{f_t^1, \dots, f_t^N\}$ が得られる。

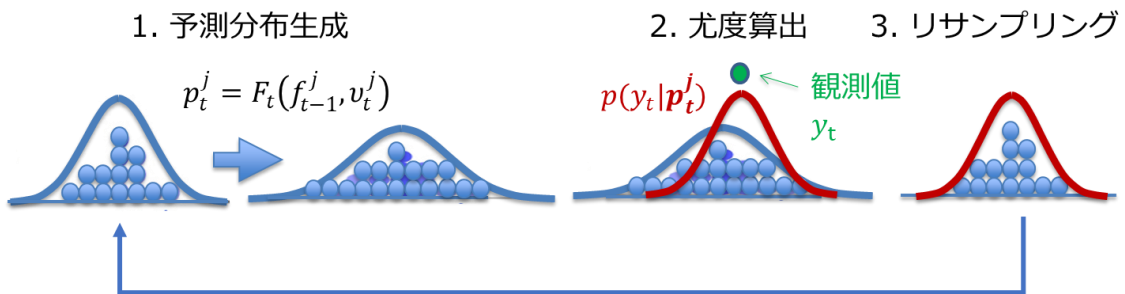


図 3.3 粒子フィルタの模式図

粒子フィルタの拡張

我々の提案した状態空間モデルのシステムモデル式3.16-3.18、観測モデル式3.22には、ポアソン分布や一様分布などが含まれ、状態の値による場合分けなどの非線形性もある。前述の粒子フィルタはこのような状態空間モデルを解くことを可能とするため、本研究では粒子フィルタを採用した。

我々のシステムモデルでは式3.18に示したように、システムノイズに幅の広い一様分布を導入して状態の大きな非定常変化に対応させている。一方、一様分布の幅を大きく広げすぎると、粒子の密度を低下させモンテカルロ誤差を増加させることに繋がる。設定した一様分布の幅を超えるような大きな非連続的な変化が発生した際に、どのようにしてパラメータ推定するか検討する必要がある。

我々は、次のような非連続性の検知と補正の方法を提案する。まず観測値が予測分布の端の値より $N\sigma$ 以上離れているかどうかを評価する。 N はハイパーパラメータであり、目的に合わせて非連続性に対する感度を調整する。ここでは $N=1$ とする。観測値が予測分布の端より 1σ 以上離れていた場合には、非連続的な観測値が生じたと見做し、粒子分布を観測値 y_t をもとに式3.16-3.18で初期化する。なお、粒子分布初期化の際には、 y_t より 1σ 小さい値を用いて式3.16-3.18で初期分布を求めることで、オーバーフィッティングを抑制する。

本研究では粒子数は10,000個とし、初期の粒子分布は $x_{-1} = y_0$ として式3.16-3.18により求めた。各時刻の平均値 λ は、各時刻の粒子分布の中央値をとって求めた。平均でなく中央値としたのは、粒子分布のばらつきに左右されにくい平均値 λ の推定を行うためである。

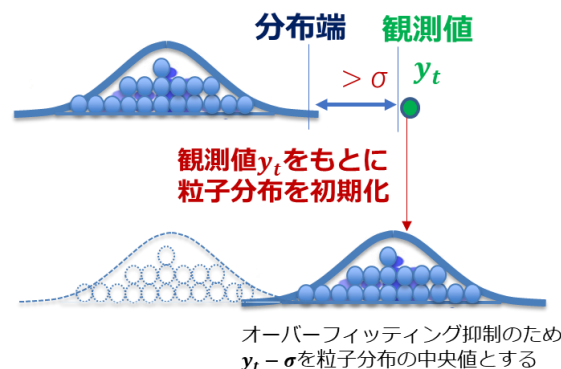


図 3.4 非連続的な観測値の検知と粒子分布の補正の模式図

提案した粒子フィルタのアルゴリズムのまとめ

提案した粒子フィルタのアルゴリズムをまとめると次のようになる。

- a) 粒子 N 個の初期分布を $x_{-1} = y_0$ として式3.16-3.18により生成し、時間 $t = 1$ とする。
 - b) 時間 t について、
 - b-1) 時間 t での粒子の予測分布を式3.16-3.18により求める
 - b-2) 観測値 y_t をもとに非連続検知を行う。 y_t が予測分布の上限 $+\sigma$ を上回る、または下限 $-\sigma$ を下回るかを評価し、当てはまるなら下記 b-3、当てはまらないなら b-3' に進む。
 - b-3) 非連続が検知されたため、粒子分布の初期化を行う。 y_t が予測分布の上限を上回る場合は $y_t - \sigma$ 、下限を下回る場合は $y_t + \sigma$ を x_t として、式3.16-3.18により粒子分布を生成する。
 - b-3') 観測値 y_t に対する各粒子の尤度を式3.22で算出し、尤度に比例した確率での粒子の復元抽出を行って、粒子分布を更新する。
 - b-4) t が最後の時間なら処理を終え、そうでない場合 t を 1 増加させ上記 b-1 に戻る。
- 図 3.5 にフローチャートを示す。

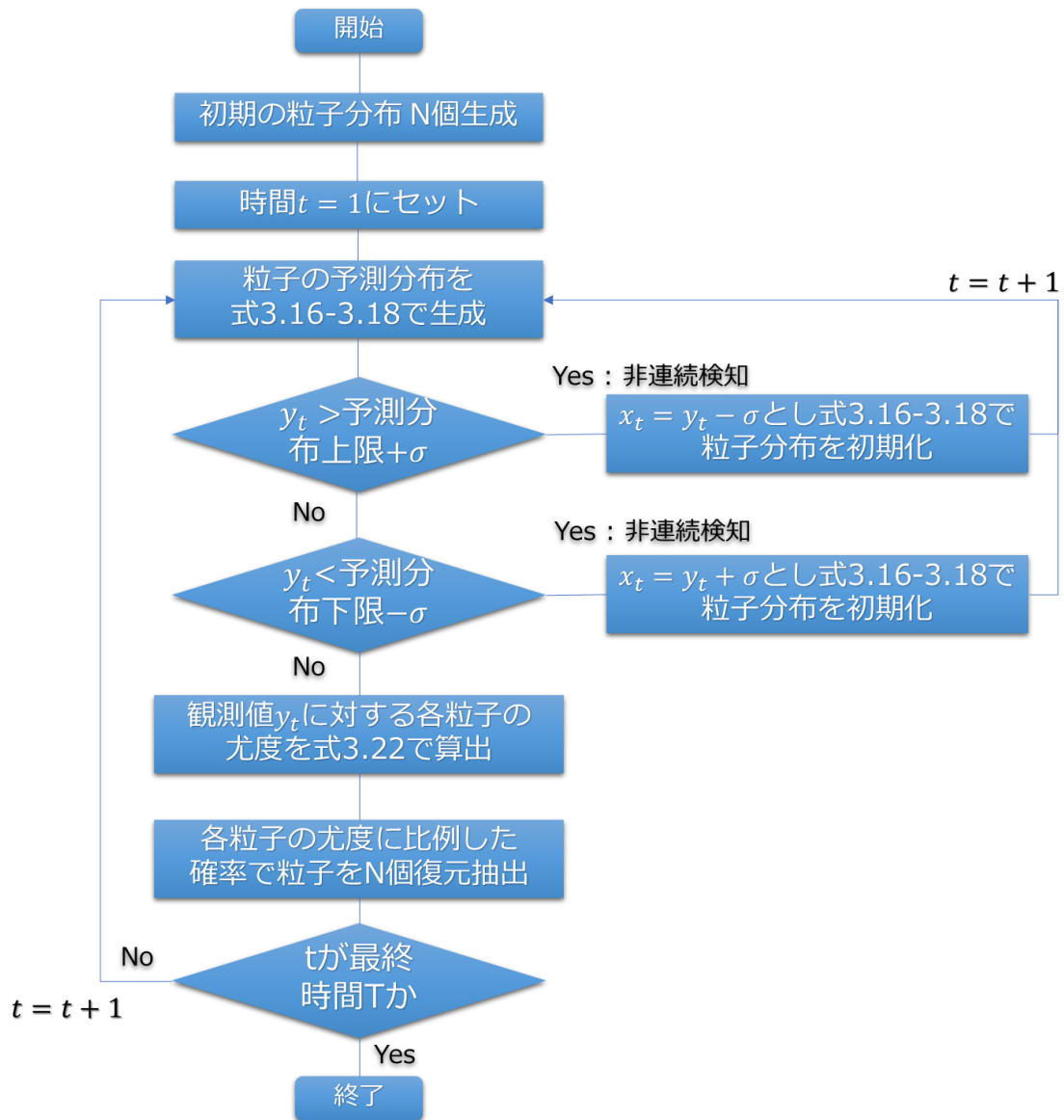


図 3.5 提案した粒子フィルタのアルゴリズム

アルゴリズムの詳細

システムモデル式3.16-3.18による予測分布の生成は次の手順で実施する。

- a) $t - 1$ 時点での状態 x_{t-1} が N 個の粒子、つまりモンテカルロ乱数で表現されているとする。
- b) 式3.18に対応したノイズ v_t を粒子数と同じ N 個発生させる。具体的には、 $-\beta$ から β の値をとる一様分布 $U(-\beta, \beta)$ からの乱数を $m \times N$ 個、平均0で標準偏差 $\alpha \cdot x_{t-1}$ の正規分布 $N(0, \alpha \cdot x_{t-1})$ からの乱数を $(1 - m) \times N$ 個の生成する。
- c) 1番目から N 番目まで、a)の粒子にb)の乱数ノイズを加えて、 t 時点での N 個の粒子の予測分布を得る。もし粒子の値が負となった場合は、式3.16-3.17に示したようにその粒子を除外する。

観測モデル式3.22による、観測値をもとにした粒子分布のフィルタリングは以下の手順で行う。

- a) t 時点での N 個の粒子それぞれの値について、観測値 y_t が得られる尤度を計算する。式3.22に従って、各粒子の値が20未満の場合はポアソン分布のPDF、20以上の場合は、標準偏差に式3.2のテイラースケール項が組み込まれた正規分布のPDFを用い、各粒子の値を x_t とした尤度を計算する。
- b) 各粒子の尤度を全粒子の尤度の和で割って1に規格化する。
- c) 各粒子のリサンプリングを行う。具体的には、各粒子の選択確率をb)で得た規格化された各粒子の尤度として、 N 個の粒子を復元抽出する。

上記の復元抽出においては各要素の選択確率を個別に設定する必要がある。このような選択確率の重み付きの復元抽出のアルゴリズムとして、Alias method [125]がある。統計分析ソフトウェアRでは、sample関数 [126]で選択確率の重み付きの復元抽出を実施できる。

ハイパーパラメータの決定

式3.18のシステムモデルのノイズ v_t は、 α, β, m の3つのハイパーパラメータを含んでいる。これらのパラメータは次のように決めることができる。まず系に典型的な時系列を想定する。例えば、図 3.6(a)の定常時系列、図 3.7(a)の階段状の時系列、図 3.8(a)の連続上昇する時系列である。乱数によりこれらの時系列を多数、例えば 100 セットずつ用意する。 α, β, m を条件振りしながら粒子フィルタで各時系列における λ の推定値を求める。乱数時系列の真の λ と推定された λ から Root Mean Squared Error (RMSE)を求め、RMSE が小さくなる α, β, m を採用する。図 3.6, 図 3.7, 図 3.8 の(b),(c),(d)は、それぞれ $m = 0.05, 0.10, 0.15$ として、 α と β を条件振りして得た RMSE のマップである。図 3.6 より、定常時系列に対しては、 α, β, m が小さい場合、つまりノイズ v_t が小さく予測分布の幅が狭い場合に、良好な λ 推定結果が得られることが分かる。これは時系列の揺らぎに鈍感となるためである。一方、図 3.7, 図 3.8 より、階段状や連続上昇といった非定常時系列に対しては、定常時系列の時とは逆の傾向となり、ノイズ v_t が大きく予測分布の幅が広いほど良好な λ 推定結果となる。広すぎる予測分布幅は推定精度低下につながっていることも分かる。これは粒子フィルタの粒子数が限られているため、広すぎる予測分布幅はモンテカルロ誤差を大きくするためと考えられる。これらの結果より、本研究では 3 つの時系列に対しバランスよく小さな RMSE を与える条件 $m = 0.05, \alpha = 0.005, \beta = 2.5\sigma_{t-1}$ を採用した。

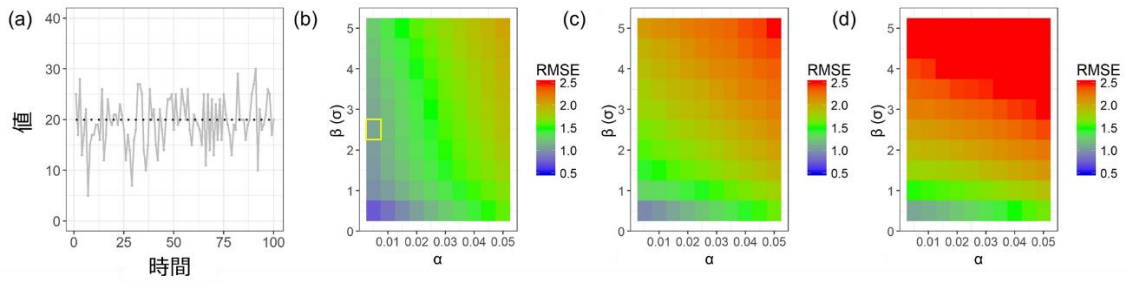


図 3.6 定常時系列による RMSE 見積り。(a)の灰色線は乱数時系列。黒点線は真の λ 。(b)は $m = 0.05$ 、(c)は $m = 0.10$ 、(d)は $m = 0.15$ で、それぞれ α, β を振った時の推定 λ と真の λ との RMSE 見積り結果。(b)中の黄色枠は採用した条件。

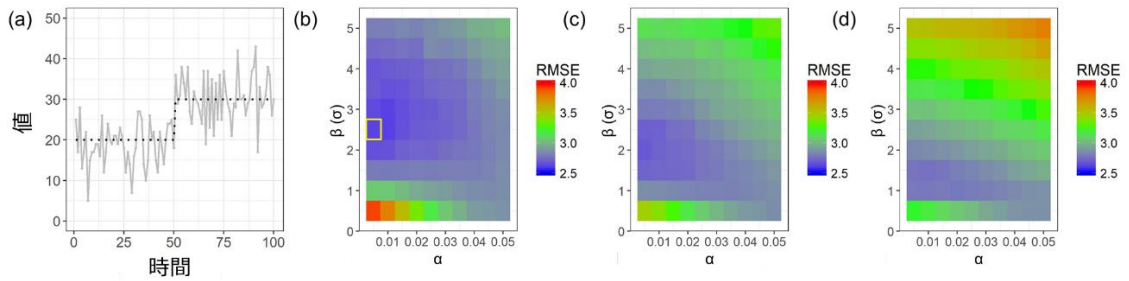


図 3.7 階段状の非定常時系列による RMSE 見積り。(a)の灰色線は乱数時系列。黒点線は真の λ 。(b)は $m = 0.05$ 、(c)は $m = 0.10$ 、(d)は $m = 0.15$ で、それぞれ α, β を振った時の推定 λ と真の λ との RMSE 見積り結果。(b)中の黄色枠は採用した条件。

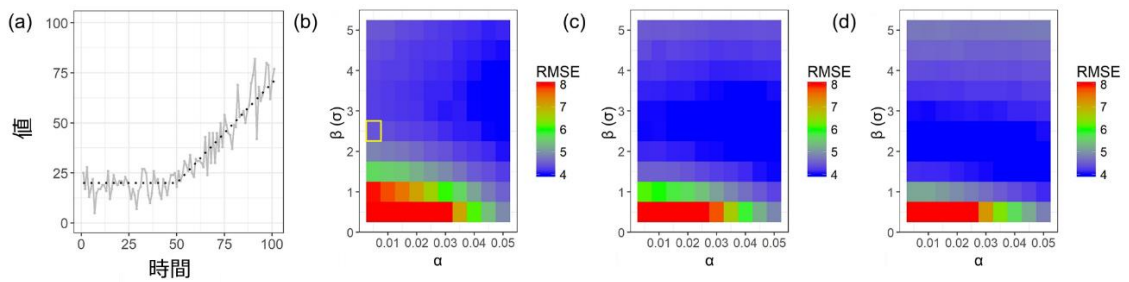


図 3.8 連続上昇の非定常時系列による RMSE 見積り。(a)の灰色線は乱数時系列。黒点線は真の λ 。(b)は $m = 0.05$ 、(c)は $m = 0.10$ 、(d)は $m = 0.15$ で、それぞれ α, β を振った時の推定 λ と真の λ との RMSE 見積り結果。(b)中の黄色枠は採用した条件。

3.4 効果検証

シミュレーション検証

本節では、乱数で人工的に生成した時系列を用いて手法の検証を行う。テイラースケーリング則に従う乱数時系列は、式3.22を用いて生成した。具体的には、乱数の平均値 λ が20未満の場合は、乱数はポアソン分布に従うとして逆関数法 [127]を用いて乱数生成した。一方、乱数の平均値 λ が20以上の場合は、乱数は標準偏差に式3.2のテイラースケーリング項を組み込んだ正規分布に従うとし、Box-Muller法 [128]で乱数生成した。なお、テイラースケーリングの比例定数 γ は本研究で用いる POS データの典型的な値である 0.1 とした [28]。

始めに非定常性、非連続性への対応を検証する。図 3.9(a)は、平均値 λ を20から200に連続的に上昇させながら生成した乱数時系列とその λ 推定結果である。灰色線は乱数時系列、黒点は真の λ 、黒線は推定 λ である。連続的な真の λ の上昇に良好に追従した λ 推定が出来ることが確認できる。真の λ を λ_A 、推定 λ を λ_E として下記の式3.33で $RMSE_\lambda$ を求めると7.74%であった。

$$RMSE_\lambda = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(1 - \frac{\lambda_{E_t}}{\lambda_{A_t}}\right)^2} \quad (3.33)$$

図 3.9(b)は階段状に非連続的に平均値 λ を20から200に上昇させた場合の乱数時系列とその λ 推定結果である。ここでは前節で述べた非連続補正を行わずに λ を求めている。黒点で表した真の λ に対し、黒線の推定 λ は遅れを伴いながら追従していることが分かる。具体的には、真の λ は時間50で λ が20から200に変化しているものの、推定 λ は時間80程度で200に到達している。 $RMSE_\lambda$ を求めると14.59%であった。この推定 λ の遅れは、式3.18で定義される粒子フィルタの予測分布の幅が、20から200に変化する非連続性より小さいことにより発生する。図 3.10 は非連続補正を導入した場合の λ 推定結果である。遅れのない推定が出来ており $RMSE_\lambda$ は6.55%に改善された。

続いてテイラースケーリング則への対応を検証する。図 3.11(a)は、真の λ を10から640まで時間30毎に2倍にして広い範囲で変化させ、標準偏差がテイラースケーリング則に従って大きく変化している時の λ 推定結果である。灰色線は乱数時系列であり、平均が大きくなるに伴い、平均に比例して揺らぎが大きくなっている。黒線の λ 推定値は、黒点の真の λ に

良好に追従しており $RMSE_\lambda$ は 8.66% であった。

ここで、推定 λ がテイラースケーリングに対して妥当なものになっているかをより詳細に評価する。推定 λ が妥当なものであれば、この推定 λ に対する観測値の標準偏差は式3.2に従うはずである。推定 λ に対する観測値の標準偏差は、次の手順で求めた。

- a) 各時点 t での推定 λ と観測値の組を作る。
- b) a) で得た各組を、各組の λ の値を基準に bin 分けする。具体的には 2 のべき乗、つまり $[1, 2), [2, 2^2), [2^2, 2^3), \dots$ の幅の bin を考え、各組を λ の値で bin 分けする。
- c) 各 bin 中の推定 λ と観測値から標準偏差を求める。なお、平均は推定 λ の平均とする。

図 3.11(b) は上記手順で得られた平均と標準偏差の関係である。黒点は bin 毎の平均と標準偏差であり、黒の実線で示したテイラースケーリングの関係に良好に追従していることが分かる。各 bin の標準偏差を σ_{E_t} とし、対応する黒の実線上の標準偏差を σ_{A_t} とし、下記の式3.34により $RMSE_\sigma$ を求めると 8.71% であった。

$$RMSE_\sigma = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(1 - \frac{\sigma_{E_t}}{\sigma_{A_t}} \right)^2} \quad (3.34)$$

従来モデルとの比較のため、図 3.11(a) と同じ乱数時系列に対して、粒子フィルタの観測モデルを、テイラースケーリングを考慮した式3.22でなく、通常のポアソン分布である式3.21とした時の λ 推定結果を図 3.11(c) に示す。黒点線の真の λ に対し、黒実線の推定 λ は揺らぎに過剰に追従していることが分かる。このことは図 3.11 (d) の推定 λ に対する観測値の標準偏差からもわかる。ポアソン分布を観測モデルとする従来ポアソンモデルでは、標準偏差の大きさは平均の平方根であると見做される。従って、テイラースケーリングによってポアソン分布から大きく乖離した揺らぎが λ の非定常変化と見做されるため、正しい λ 推定が行われず、推定 λ は揺らぎに過度に追従する結果となる。式3.34により求めた $RMSE_\sigma$ は 33.73% であった。

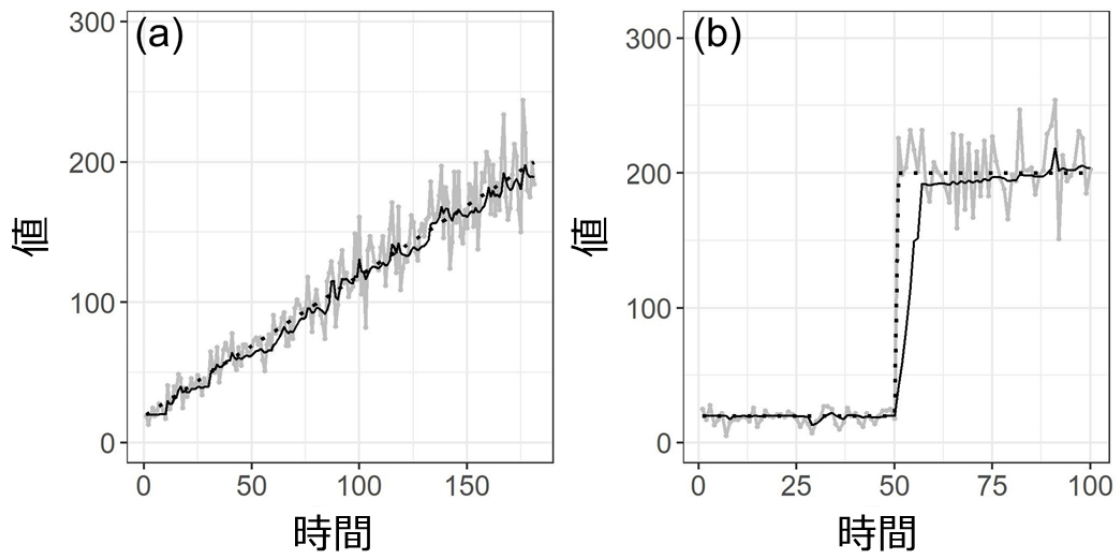


図 3.9 乱数で生成した 2 種類の非定常時系列の λ 推定結果。(a) 平均値 λ を 20 から 200 に連続的に上昇させて生成した乱数時系列、(b) 平均値 λ を 20 から 200 に階段状に変化させて生成した乱数時系列。灰色線は乱数時系列、黒点は真の λ 、黒線は推定 λ 。

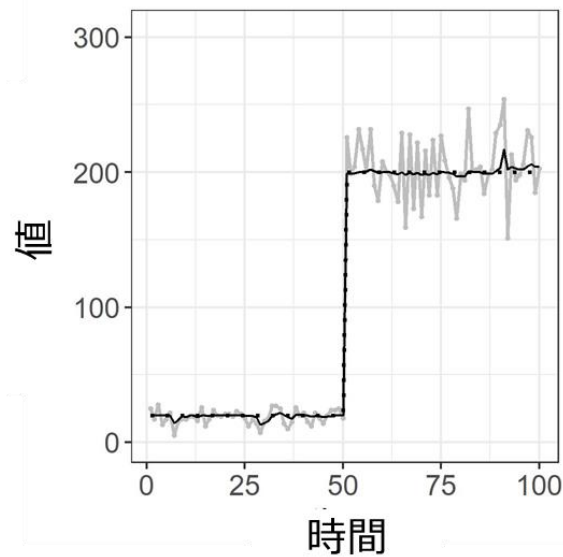


図 3.10 非連続補正による推定 λ の追従性改善。灰色線は乱数時系列、黒点は真の λ 、黒線は推定 λ 。

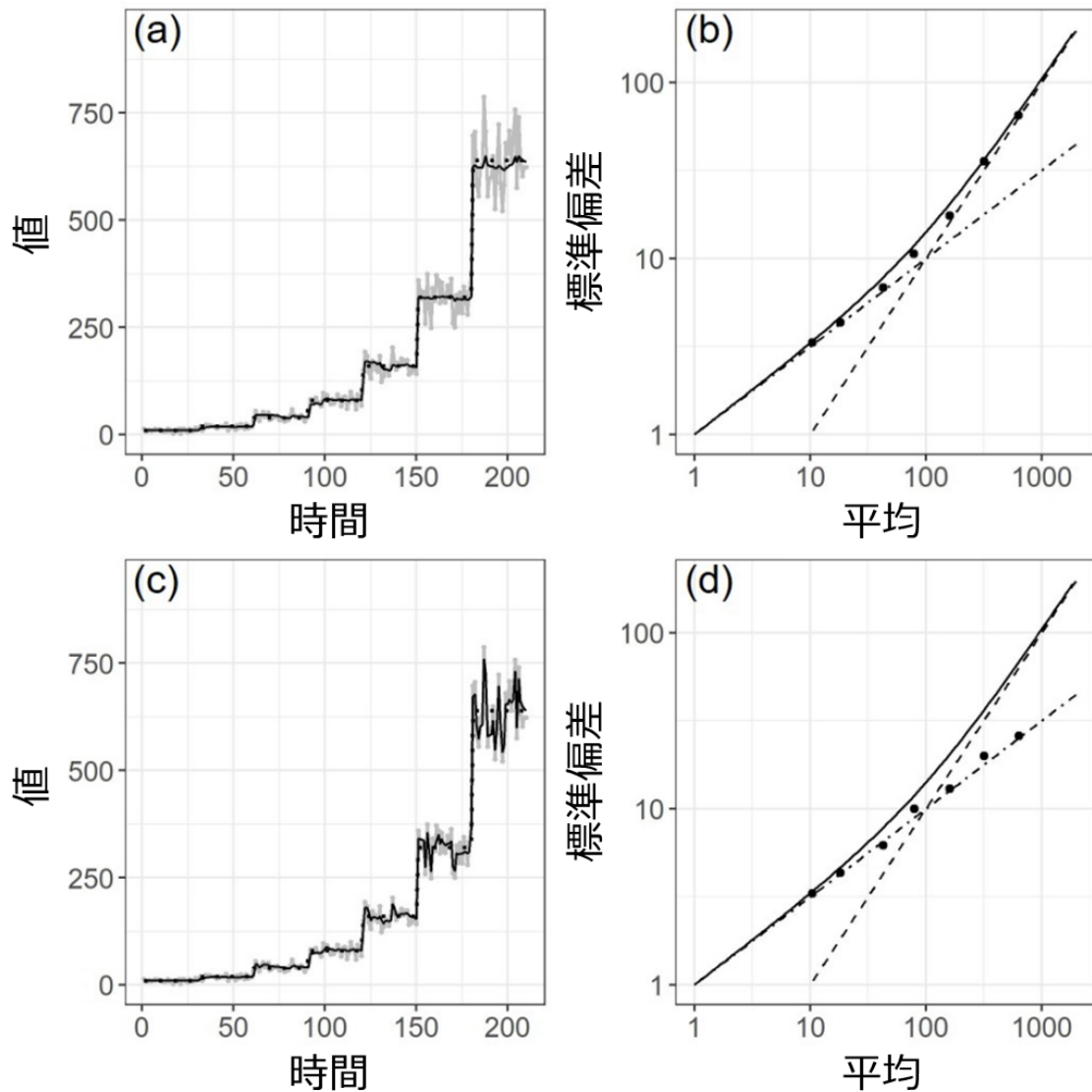


図 3.11 テイラースケーリングへの対応性の評価。(a) 乱数時系列の真の λ を 10 から 640 まで時間 30 毎に 2 倍に変化させた場合の λ 推定結果。灰色線は乱数時系列、点線は真の λ 、黒実線は推定 λ 。(b) 推定 λ に対する観測値の標準偏差。黒点は 2 のべき乗で推定 λ を Bin 分けし各 Bin の中で求めた平均と標準偏差、黒実線はテイラースケーリングの平均と標準偏差、黒点線は平均に比例した標準偏差、黒一点鎖線は平均の平方根とした標準偏差。(c) 通常のポアソン分布を観測モデルとして(a)と同じ乱数時系列の λ 推定を行った結果。(d) (c)の推定 λ に対する観測値の標準偏差。

POS データによる検証

POS 販売時系列を用いて手法の検証を行った。図 3.12(a)は、ある店舗のあるアイスクリームの販売時系列である。灰色線は販売時系列であり、大きな上昇下降がみられる。黒の実線で示した λ 推定結果はこの上昇下降に良好に追従していることが確認される。図 3.12(b)は推定 λ と観測値による平均と標準偏差の関係である。推定 λ と観測値から得た黒点の平均と標準偏差は、黒実線で表したテイラースケーリングの関係と整合していることが分かる。 $RMSE_{\sigma}$ は 16.56%であった。

図 3.12(c),(d)は、図 3.12(a)と同様な時系列をポアソンモデルで λ 推定した結果である。図 3.11(c),(d)で見られたのと同様に揺らぎに過度に追従していることが分かる。 $RMSE_{\sigma}$ は 35.40%であった。

図 3.13(a)はある店舗の肉まんの販売時系列である。灰色線は販売時系列であり、時間 10 と 50 付近に非連続的な大きな上昇下降があることが分かる。この時系列においても、黒線で示した本手法の推定 λ は変化に良好に追従している。図 3.13(b)から、この λ 推定結果はテイラースケーリングの関係と整合していることが確認できる。 $RMSE_{\sigma}$ は 18.74%であった。図 3.13(c),(d)は提案手法において、非連続補正を実施しない場合の λ 推定結果である。推定 λ は非連続な上昇に追従できておらず、 $RMSE_{\sigma}$ は 337.62%であった。非連続補正の有効性が実際の POS 販売時系列からも確認された。

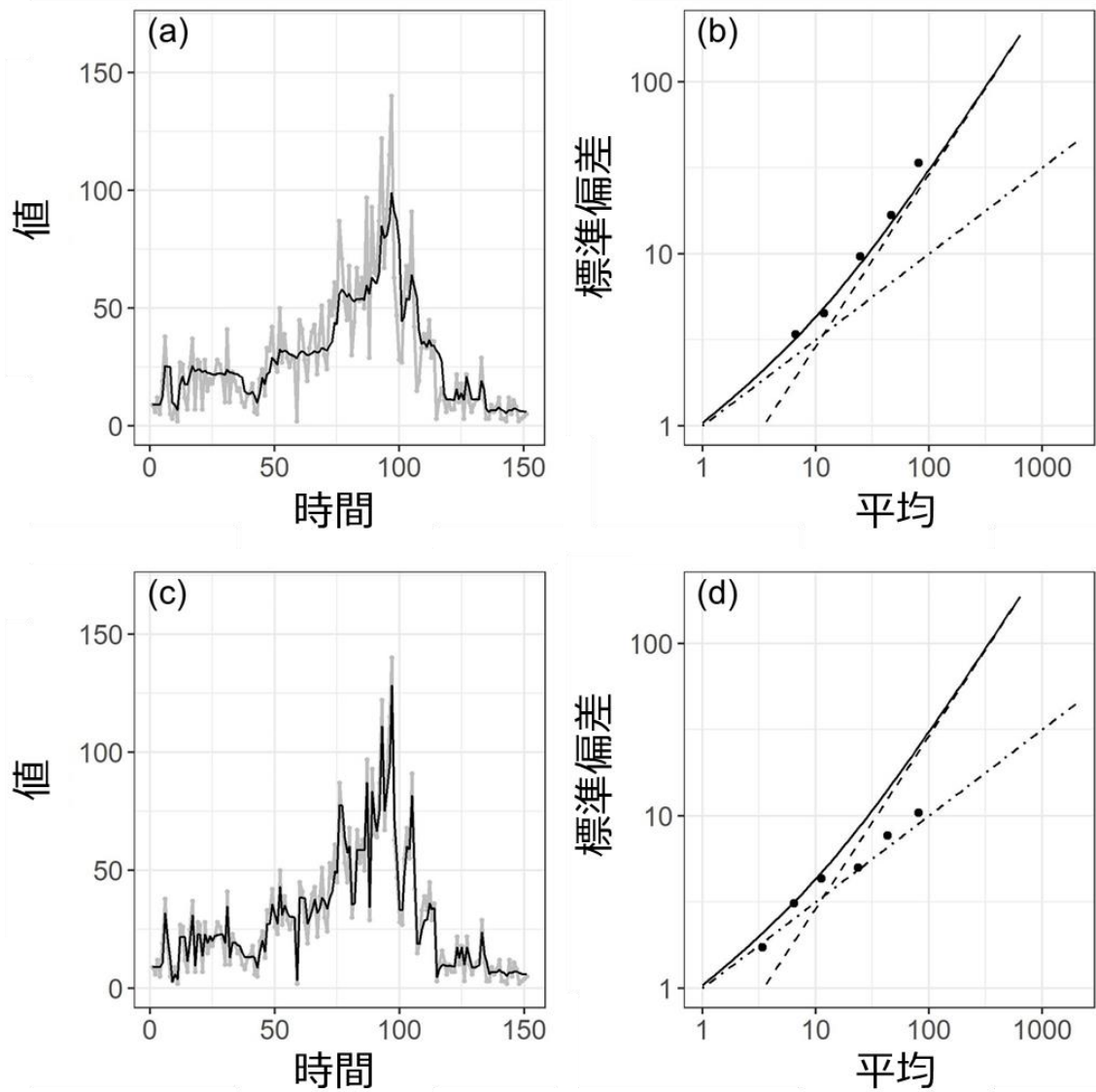


図 3.12 上昇下降がある POS データでの λ 推定結果。(a) 提案手法による λ 推定結果。灰色線は POS 販売時系列、黒線は λ 推定結果。(b) 推定 λ と観測値による平均と標準偏差の関係。黒点は 2 のべき乗で推定 λ を Bin 分けして求めた平均と標準偏差、黒実線はテイラースケーリングの関係、黒点線は平均に比例した標準偏差、黒一点鎖線は平均の平方根に比例した標準偏差。(c) (a) と同じ時系列をポアソンモデルで λ 推定した結果。(d) (c) の推定 λ と観測値による平均と標準偏差の関係。テイラースケーリングの比例定数 γ は 0.291 である。

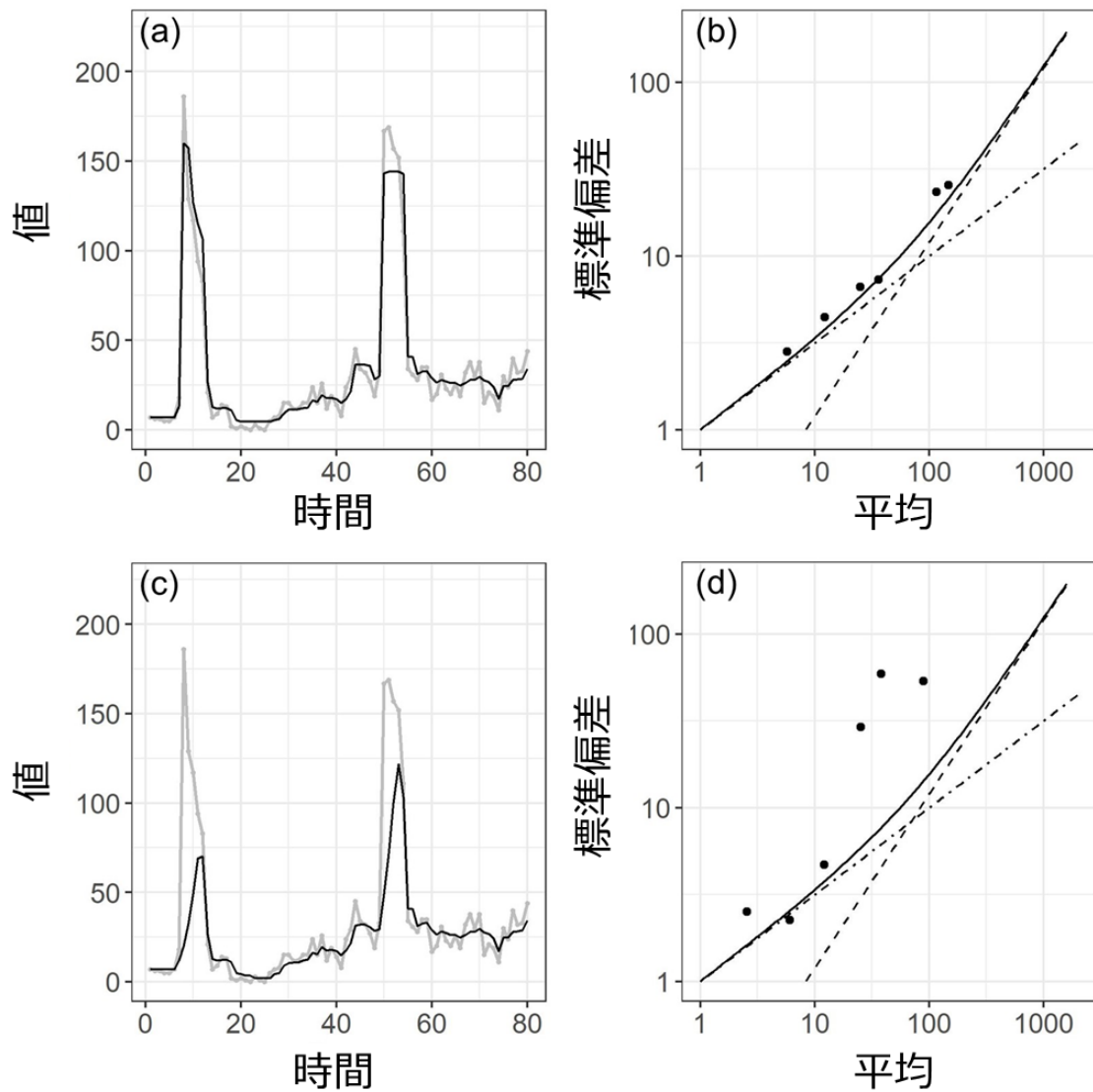


図 3.13 非連続的な大きな変化がある POS データの λ 推定結果。(a) 提案手法による λ 推定結果。灰色線は POS の販売数、黒線は λ 推定結果。(b) 推定 λ と観測値による平均と標準偏差の関係。黒点は 2 のべき乗で Bin 分けして求めた平均と標準偏差、黒実線はテイラー・スケーリングの関係、黒点線は平均に比例した標準偏差、黒一点鎖線は平均の平方根に比例した標準偏差。(c) (a) と同じ時系列をポアソンモデルで λ 推定した結果。(d) (c) の推定 λ と観測値による平均と標準偏差の関係。テイラー・スケーリングの比例定数 γ は 0.12 である。

3.5 まとめ

本章では、販売データの統計的性質であるテイラースケーリング則を考慮した非定常ポアソン時系列のパラメータ推定手法を確立した。非定常時系列分析手法である粒子フィルタの観測モデルにおいてテイラースケーリング則を考慮した尤度関数を導入した。併せて、非連続的な時系列変化に対応するための非連続検知と補正の手法を開発し粒子フィルタに導入した。提案手法の効果を、乱数で人工的に生成した時系列とPOSの販売時系列で検証し、テイラースケーリング則による標準偏差のポアソン過程からの乖離が存在し、非連続的な変化を含む非定常な販売時系列において、提案手法によってデータと整合する適切なパラメータ推定に可能となることを確認した。

第4章 非定常時系列分析手法の廃棄低減 への応用

第4章では、第3章で確立した非定常時系列分析手法を応用し、消費期限のある商品に対し、廃棄を低減しつつ利益の減少を最低限に留めるように仕入数を決定する手法を提案する。まず、仕入数が需要に満たない時には売り切れにより、観測される販売数は潜在的な需要分布の下限側の値となることを指摘し、第3章で提案した粒子フィルタの観測モデルに、観測上限で打ち切られたデータに対する既知の最尤推定法を導入する。さらに機会損失と廃棄損失の最適化問題の解として知られる新聞売り子問題の公式を拡張し、利益ロスを見積もりながら所望の廃棄率を実現する仕入数を決定できる新しい手法を提案する。提案手法の効果を検証するため、店舗毎、商品毎のPOS販売時系列をそれぞれ各店舗、各商品の日々の需要と見做した販売シミュレーションを行う。さらに、複数店舗、複数商品のPOSデータを集計した販売時系列を用意して、販売数平均が大きくテイラースケーリングの効果が明瞭となる販売時系列に対する効果検証を行う。なお、本章の内容は論文 [129, 130] に基づく。

4.1 観測打ち切りへの対応

消費期限のある商品においては、小売店は一般に過剰な在庫をもたず、想定される需要に対して、売れ残りによる廃棄損失と売り切れによる機会損失を最小化する仕入を行う。この仕入数が実際の需要に満たない場合には売り切れが発生する。この時、需要の観測値である販売数は、潜在的な需要の分布の下限側の値となる。図4.1はその模式図である。図4.1(a)は販売時系列の例であり、灰色線は潜在需要、橙線は仕入数、緑線は販売数、赤線は潜在需要の平均値を表す。図4.1(b)は、図4.1(a)中の肌色の四角領域の拡大図である。橙線の仕入数が灰色線の潜在需要より小さい時、観測される緑線の販売数は潜在需要より小さくなる。図4.1(c)は、潜在需要の分布関数と、観測された需要の関係の模式図である。需要の観測が仕入数で打ち切られることで、潜在需要分布は下限側の値しか観測されない。売り切れがある場合の潜在需要平均の推定は、このような打ち切られた観測需要から行うことが必要となる。本節では、前章で提案した粒子フィルタを拡張し、観測上限で打ち切られた非定常な販売時系列からのパラメータ推定を可能とする。

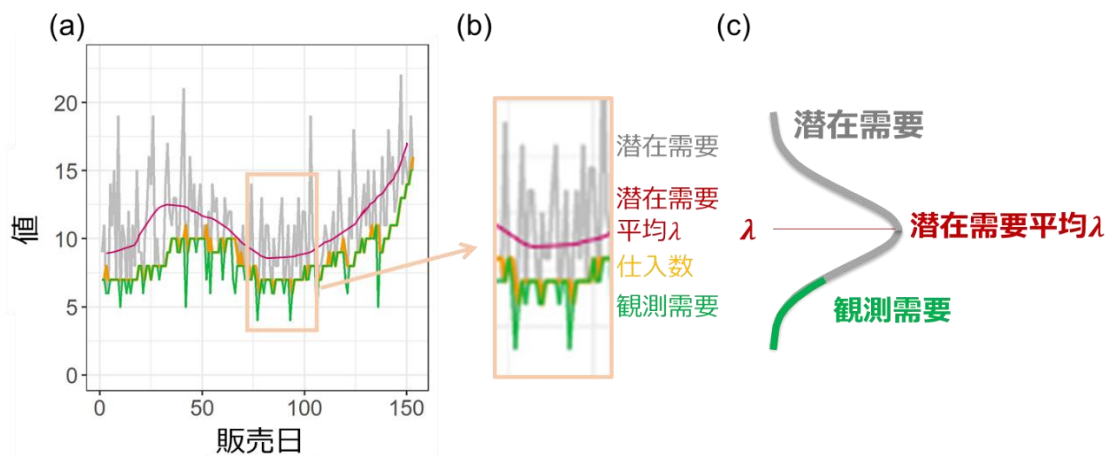


図 4.1 仕入数による販売数の観測打ち切り。(a) 販売時系列の例。灰色線は潜在需要、橙線は仕入数、緑線は販売数、赤線は潜在需要の平均値。(b) 左図中の肌色の四角領域の拡大図。(c) 潜在需要の分布関数と観測された需要の関係の模式図。

前章の粒子フィルタでは、観測値に対する各粒子の尤度を式3.22により求めた。ここでは、時間 t での需要が仕入数 s_t より大きく、観測される販売数が観測上限 s_t で打ち切られた場合の尤度を考える。 s_t が観測される確率は s_t 以上の値が観測される確率の和となる。そのため、図4.2に示したように尤度関数は次の式で与えられることが、打ち切りデータの最尤推定法として確立されている [131]。

$$F(s_t|\theta) = \int_{s_t}^{\infty} f(k|\theta) dk \quad (4.1)$$

$f(k|\theta)$ はパラメータ θ の下で k が得られる確率密度関数(PDF)であり、尤度関数 $F(s_t|\theta)$ は累積分布関数(CDF)に対応している。本研究では、次のようなテイラースケーリングを考慮した打ち切りデータの最尤推定法を導入した。

$$p(y_t|x_t) = \begin{cases} \frac{\exp\left(\frac{-(y_t - x_t)^2}{2\sigma_t^2}\right)}{\sqrt{2\pi\sigma_t^2}}, & x_t \geq 20 \text{ で打ち切りなし} \\ \frac{\int_{y_t}^{\infty} \exp\left(\frac{-(m - x_t)^2}{2\sigma_t^2}\right) dm}{\sqrt{2\pi\sigma_t^2}}, & x_t \geq 20 \text{ で打ち切りあり} \\ \frac{x_t^{y_t} \cdot \exp(-x_t)}{y_t!}, & x_t < 20 \text{ で打ち切りなし} \\ \sum_{m=y_t}^{\infty} \frac{x_t^m \cdot \exp(-x_t)}{m!}, & x_t < 20 \text{ で打ち切りあり} \end{cases} \quad (4.2)$$

ここで σ_t はテイラースケーリング則に対応した標準偏差 $\sigma_t = \sqrt{x_t + (\gamma \cdot x_t)^2}$ である。観測打ち切りがない時には式3.22と同様であり、観測が y_t で打ち切られた時には $x_t \geq 20$ の時には σ_t を標準偏差とする正規分布、 $x_t < 20$ の時はポアソン分布を確率分布として、それぞれ y_t 以上になる確率を足し合わせたものを尤度とする。

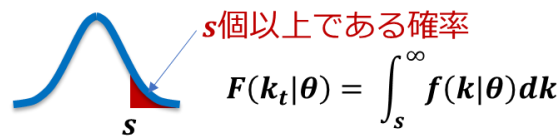


図 4.2 観測上限 s で打ち切られた時の尤度関数 $F(k_t|\theta)$ 。 k_t は観測値、 θ はパラメータ、 $f(\cdot|\theta)$ は任意のPDF。

4.2 新聞売り子問題とその拡張

新聞売り子問題

前節の提案手法により、売り切れを含む不完全な販売データから需要平均 λ を推定できる。一方、需要平均 λ の知見を店舗での利益向上や廃棄低減に活用するためには、 λ から仕入数をどのように決めるかが課題となる。需要は一定値でなく平均 λ の周りで確率密度関数 $m(k|\lambda)$ に従って揺らぐと考える時に、どのような仕入数が最適かは自明でなく、機会損失と廃棄損失の最適化問題を解く必要がある。新聞売り子問題の公式は、その最適化問題の解として知られている。需要分布が $m(k|\lambda)$ 、商品の原価が c 、売値が p 、仕入数が s の時に、期待される利益 $R(s)$ は、

$$R(s) = p \left\{ \sum_{k=0}^s km(k|\lambda) + \sum_{k=s+1}^{\infty} sm(k|\lambda) \right\} - cs \quad (4.3)$$

最大利益となる仕入数を求めるため、式4.3の差分を0とおくと、

$$R(s+1) - R(s) = p \sum_{k=s+1}^{\infty} m(k|\lambda) - c = 0$$

これより、最大利益となる仕入数 s^* は $m(k|\lambda)$ の累積分布関数の逆関数 $M^{-1}(k)$ を用いて次のように書ける [68]。

$$s^* = M^{-1}\left(\frac{c}{p} \mid \lambda\right) \quad (4.4)$$

この式4.4は、需要分布 $m(k|\lambda)$ の累積分布関数の逆関数 $M^{-1}(k)$ と原価率 c/p によって、最大利益となる仕入数 s^* が決まることを意味する。

なお、提案手法では $m(k|\lambda)$ はテイラースケーリングの揺らぎを考慮した次の式4.5に従うと仮定する。

$$m(k|\lambda) = \begin{cases} Po(\lambda), & \text{if } \lambda < 20 \\ N\left(\lambda, \sqrt{\lambda + (\gamma \cdot \lambda)^2}\right), & \text{otherwise} \end{cases} \quad (4.5)$$

図4.3(a)は、需要平均10、原価率0.7の時の機会損失、廃棄損失と仕入数の関係である。赤線は機会損失、緑線は廃棄損失、青線の機会損失と廃棄損失の和である。機会損失は式4.6、廃棄損失は式4.7のように書ける。

$$(p - c) \sum_{k=s+1}^{\infty} (k - s)m(k|\lambda) \quad (4.6)$$

$$c \sum_{k=0}^s (s-k)m(k|\lambda) \quad (4.7)$$

機会損失と廃棄損失の和は二次曲線状であり、損失の和を最小にする仕入が存在することが分かる。式4.4はこの最適仕入を求めるものである。図4.3(b)は機会損失と廃棄損失の関係である。青点と赤点は、後述する図4.8の(a),(c)の販売時系列から得た機会損失と廃棄損失である。

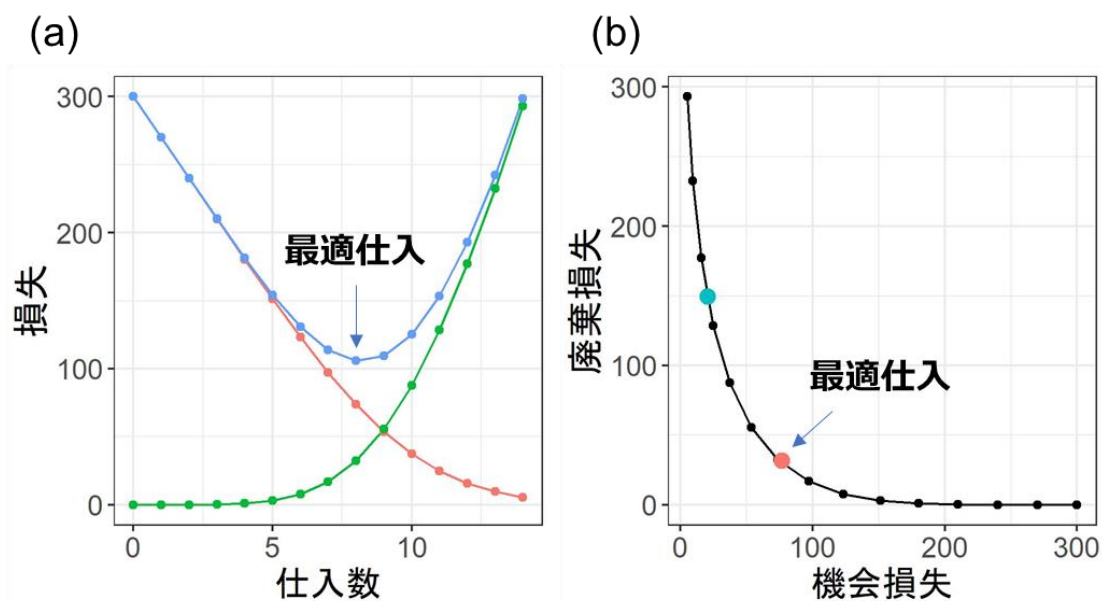


図 4.3 機会損失と廃棄損失の関係。(a) 需要平均 10、原価率 0.7 の時の機会損失、廃棄損失と仕入数の関係。赤線は機会損失、緑線は廃棄損失、青線は機会損失と廃棄損失の和。(b) 機会損失と廃棄損失の関係。青点と赤点は図 4.8 の(a)と(c)の機会損失と廃棄損失に相当。

図 4.4 は、原価率、需要平均 λ 、需要分布を変化させた時に式4.4から得られる最適仕入数である。茶線、赤線、緑線、青線はそれぞれ原価率 0.3、0.5、0.7、0.9 を表す。需要分布について実線は式4.5と仮定し、点線はポアソン分布と仮定している。テイラースケリングの比例定数 γ は 0.1 とした。原価率 0.5 の時は最適仕入数 s^* は需要平均 λ と等しい。原価率が 0.5 を超えると s^* は λ より小さく、原価率が 0.5 を下回ると s^* は λ より大きくなる。需要分布にテイラースケリングを考慮しない点線の場合の仕入数は、テイラースケリングを考慮する実線の仕入数と異なる。需要平均 100 において、原価率

0.7では2%、原価率0.9では6%のずれが生じている。実際の需要分布はテイラースケーリングに従うことから [28]、テイラースケーリングの考慮により最適仕入数を求める精度を向上させることが出来る。

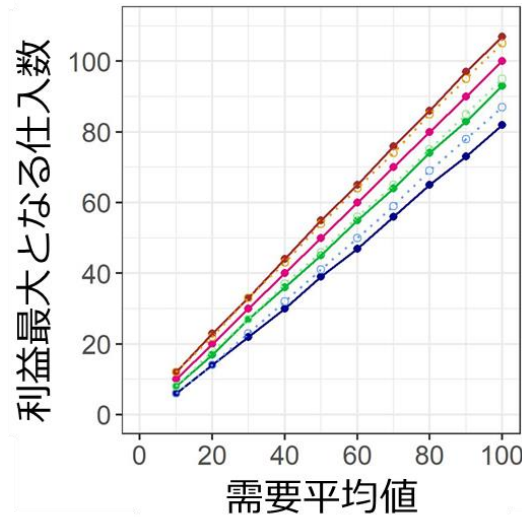


図 4.4 原価率、需要平均値、需要分布を変化させた時に新聞売り子問題の式4.4から得られる最適仕入数。茶線、赤線、緑線、青線はそれぞれ原価率 0.3、0.5、0.7、0.9。実線は需要分布を式4.5と仮定、点線は需要分布をポアソン分布と仮定。テイラースケーリングの比例定数 γ は 0.1。

新聞売り子問題の拡張

本研究では、新聞売り子問題の式4.4を廃棄低減と利益確保の両立に適したものに拡張する。最大利益となる仕入数 s^* の下で期待される廃棄数 d^* は次のように書ける。

$$d^* = \sum_{k=0}^{s^*-1} (s^* - k)m(k|\lambda) \quad (4.8)$$

従って廃棄数 d^* を α 倍($0 \leq \alpha \leq 1$)にする仕入数 $s(\alpha)$ は次の式4.9で求められる。

$$s(\alpha) = \operatorname{argmin} \left(\left| \alpha d^* - \sum_{k=0}^{s-1} (s - k)m(k|\lambda) \right| \right) \quad (4.9)$$

この仕入数 $s(\alpha)$ の時の利益は、次のように書ける。

$$R(s(\alpha)) = p(s(\alpha) - \alpha d^*) - cs(\alpha) \quad (4.10)$$

これより、廃棄数 d^* を α 倍にする時の利益低下率は $R(s(\alpha))/R(s^*)$ で求められる。

式4.9を用いて仕入数 $s(\alpha)$ を決める際には考慮すべき点がある。需要分布 $m(k|\lambda)$ は式4.5で示したように $\lambda < 20$ の時はポアソン分布である。ポアソン分布は離散分布であるため式4.9で決まる仕入数 $s(\alpha)$ は離散的になる。 $s(\alpha)$ が離散的であることは廃棄量の制御性を著しく低下させる。例えば、 $\lambda=10$ を仮定し、廃棄量を最大利益となる仕入数 s^* の時の α 倍 ($0.5 \leq \alpha \leq 1$) にしようとすることを考える。この時、式4.9から得られる仕入数 $s(\alpha)$ は2つの値7、8のみとなり、廃棄量を細かく制御することができない。そこで、実数の仕入数が得られるように、需要分布の式4.5のポアソン分布を Γ 分布を用いて連続化する。

$$m(k|\lambda) = \begin{cases} \frac{\lambda^k \cdot \exp(-\lambda)}{\Gamma(k+1)}, & \text{if } \lambda < 20 \\ N(\lambda, \sqrt{\lambda + (\gamma \cdot \lambda)^2}), & \text{otherwise} \end{cases} \quad (4.11)$$

ここで、もう一つ考慮すべき点がある。それは、所望の廃棄率 α を実現するための仕入数 $s(\alpha)$ が実数で決まったとしても、実際の小売店でのおにぎりや肉まんなどの仕入れは離散値であることである。そこで、実数の仕入数 $s(\alpha)$ が沢山の試行の結果として実現されるような、 $s(\alpha)$ を挟む2つの整数値による次のような確率的な仕入れを提案する。

$$\begin{cases} P(X = \underline{s}) = 1 - (s - \underline{s}) \\ P(X = \underline{s} + 1) = (s - \underline{s}) \end{cases} \quad (4.12)$$

ここで、 \underline{s} は仕入数 s の整数部である。この式4.12により、例えば仕入数が10.3であった時には、仕入数10を0.7の確率で、仕入数11を0.3の確率で選択し、多数の試行の結果、仕入数10.3が実現されるようにする。

需要を推定から仕入数決定までの手順

図4.5に、提案手法によって販売時系列から需要を推定し、仕入数を決定するまでの手順をまとめる。具体的な流れは次のとおりである。

1. 事前準備として、狙い廃棄率 α を、式4.9、式4.10を用いて廃棄低減コストを考慮しながら決定する。なお、テイラースケリングの比例定数 γ は3.1節で述べた方法で決定する。粒子フィルタのハイパーパラメータも3.3節で述べた方法で見積もっておく。

2. 時間 t を1とする。 $t = 1$ での仕入数 s_1 は経験的に決める必要がある。本研究では簡単のため、 $t = 1$ での販売数 k_1 と式4.9から s_1 を決定した。
3. 時間 t での販売数 k_t と廃棄数 d_t を観測する。観測された販売数 k_t が仕入数 s_t と等しい場合には、販売数の観測打ち切りが発生したと見做す。
4. 時間 t での需要平均 λ を3.3節で述べた粒子フィルタを用いて求める。粒子フィルタの観測モデルは観測打ち切りを考慮した式4.2、システムモデルは式3.16-3.18とする。上記ステップ3で得た販売数 k_t と観測打ち切り有無の情報をもとに粒子フィルタの粒子分布を更新し、得られた時間 t での粒子分布の中央値から λ を算出する。
5. 時間 t での需要分布 $m(k|\lambda)$ を式4.5により求める。具体的にはステップ4で求めた需要平均 λ 、並びにステップ1で求めたテイラースケーリングの比例定数 γ を用いて、式4.5により $m(k|\lambda)$ を決定する。
6. 時間 $t + 1$ の仕入数 $s(\alpha)$ を式4.9を用いて求める。式4.9中の最大利益となる仕入数 s^* の時の廃棄数 d^* を求めるために、ステップ5で求めた需要分布 $m(k|\lambda)$ 、及び分析対象としている商品の原価率の情報をもとに、まず式4.4から最大利益となる仕入数 s^* を求める。次に式4.8から d^* を得て、ステップ1で決めた狙い廃棄率 α に対応する仕入数 $s(\alpha)$ を式4.9から決定する。
7. 時間 $t + 1$ での仕入数 s_{t+1} を式4.12から決定する。ここでは、ステップ6で得た実数値の仕入数 $s(\alpha)$ から整数値の仕入数 s_{t+1} を決定している。
8. 時間 t を1増加させてステップ3に戻る。

上記ステップ4の粒子フィルタの具体的な計算方法は3.3節で述べた。観測値に対する各粒子の尤度の計算方法は、観測された販売数が打ち切られていない場合には3.3節と同じである。一方、観測された販売数が打ち切られていた場合には、式4.2に従って、各粒子の値が20未満の場合はポアソン分布のCDF、20以上の場合は、標準偏差に式3.2のテイラースケーリング項を組み込んだ正規分布のCDFを用いて各粒子の尤度を計算する。

ステップ6の式4.9の argmin 中の関数 $L = |\alpha d^* - \sum_{k=0}^{s-1} (s-k)m(k|\lambda)|$ の最小化は、数値計算によって実施する。具体的には、仕入数 s を変化させながら L が最小となる s を求める。

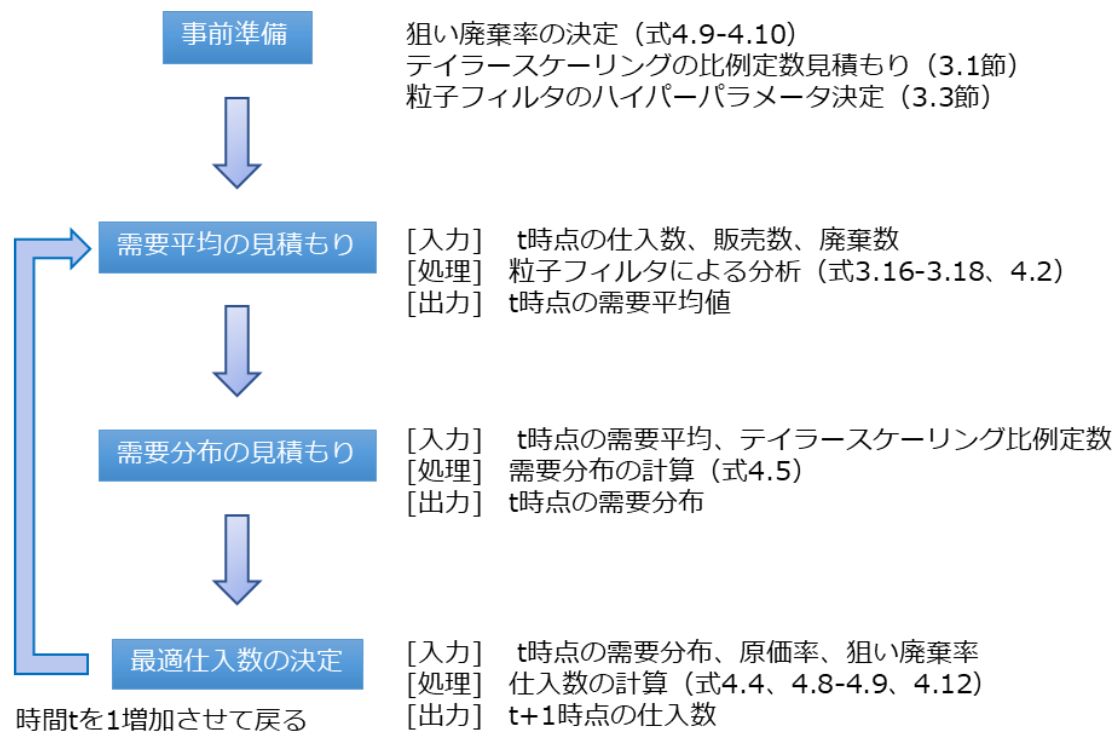


図 4.5 販売時系列から需要を推定し仕入数を決定するまでの手順

提案手法の前提条件

提案手法の前提条件をまとめる。消費期限は1日と仮定し、仕入れた商品の売れ残りは1日で廃棄される。消費期限が近付いた際の値引きは実施しないものとしている。需要推定と仕入決定をしてから商品が入荷されるまでの時間は考慮していない。商品毎の原価率は、セブンイレブンの投資家向け広報 [111]より原価率は0.7とした。予約販売や天気などの事前情報は考慮せず需要推定を行っている。

提案手法をスーパーやレストランなど様々な業態の販売データに応用する際には、業態ごとの前提条件の違いに注意する必要がある。消費期限が間近の商品の値引き販売が実施される場合には、廃棄が生じにくい。原価率が小さい場合には廃棄の利益への影響が小さくなり、提案手法の効果が得られにくくなる可能性がある。

4.3 各店舗各商品の販売時系列での効果検証

乱数時系列による検証

提案手法によって、観測打ち切りがある販売時系列でパラメータ推定ができることを、乱数時系列を用いて確認する。図 4.6(a)は真の λ を 50、テイラースケーリングの比例定数 γ を 0.1 とし、式4.5の確率分布を仮定して生成した乱数時系列である。緑線は乱数時系列であり、需要に対応する。灰色線は真の λ であり、需要平均に対応する。図 4.6(b)は図 4.6(a)の乱数時系列を潜在需要と仮定して、図 4.5 の手順により、各時間での観測需要と廃棄数から平均需要を推定し、次の時間での仕入数を決定することを繰り返して得た販売時系列である。灰色線は潜在需要、緑線は観測需要、橙線は仕入数、青線は廃棄数、赤線は推定された需要平均である。仕入数は、原価率を 0.7 とし、新聞売り子問題の式4.4で得た。緑線の観測需要は橙線の仕入数でしばしば打ち切られており、灰色線の潜在需要分布の下限側の値のみ観測されている。一方、赤線で示した推定需要平均は真の需要平均 50 程度を示している。

図 4.6(c)は仕入数決定に式4.9を用いて狙い廃棄率 $\alpha = 0.5$ とした時の販売時系列である。仕入数は図 4.6(c)の時の方が図 4.6(b)の時よりも小さくなっており、打ち切りの発生頻度が高くなっている。この条件下においても推定された需要平均は概ね 50 付近となっている。推定精度を定量化するため、真の需要平均と推定された需要平均の RMSE を計算した。図 4.6(d)は狙い廃棄率毎の RMSE である。ここでは、狙い廃棄率毎に乱数時系列を各 200 セット用意して RMSE を求め、中央値を黒点、四分位範囲をエラーバーで表示している。RMSE は狙い廃棄率が小さくなるほど大きくなっており、狙い廃棄率 1.0 では RMSE 6.6%、狙い廃棄率 0.5 では RMSE 7.5%であった。これは、狙い廃棄率が低くなるほど仕入数が小さくなり、打ち切り頻度が上昇するため、需要平均の推定が難しくなることに対応する。

続いて、非定常で観測打ち切りがある時系列に対する提案手法の有効性を検証する。図 4.7(a)は平均 50、振幅 30、周期 150 のサインカーブ状の平均値をもった乱数時系列である。緑線は観測需要、灰色線は真の需要平均である。図 4.7(b),(c)は図 4.6(b),(c)と同様に狙い廃棄率を 1.0、0.5 とし得た販売時系列である。赤線の推定需要平均は、灰色線の真の需要平均に概ね追従していることが分かる。図 4.7(d)は図 4.6(d)と同様に求めた真の需要平均と推定需要平均との RMSE である。狙い廃棄率が小さいほど RMSE が上昇する傾向がみられるものの、RMSE 10%前後で需要平均が推定できていることが分かる。

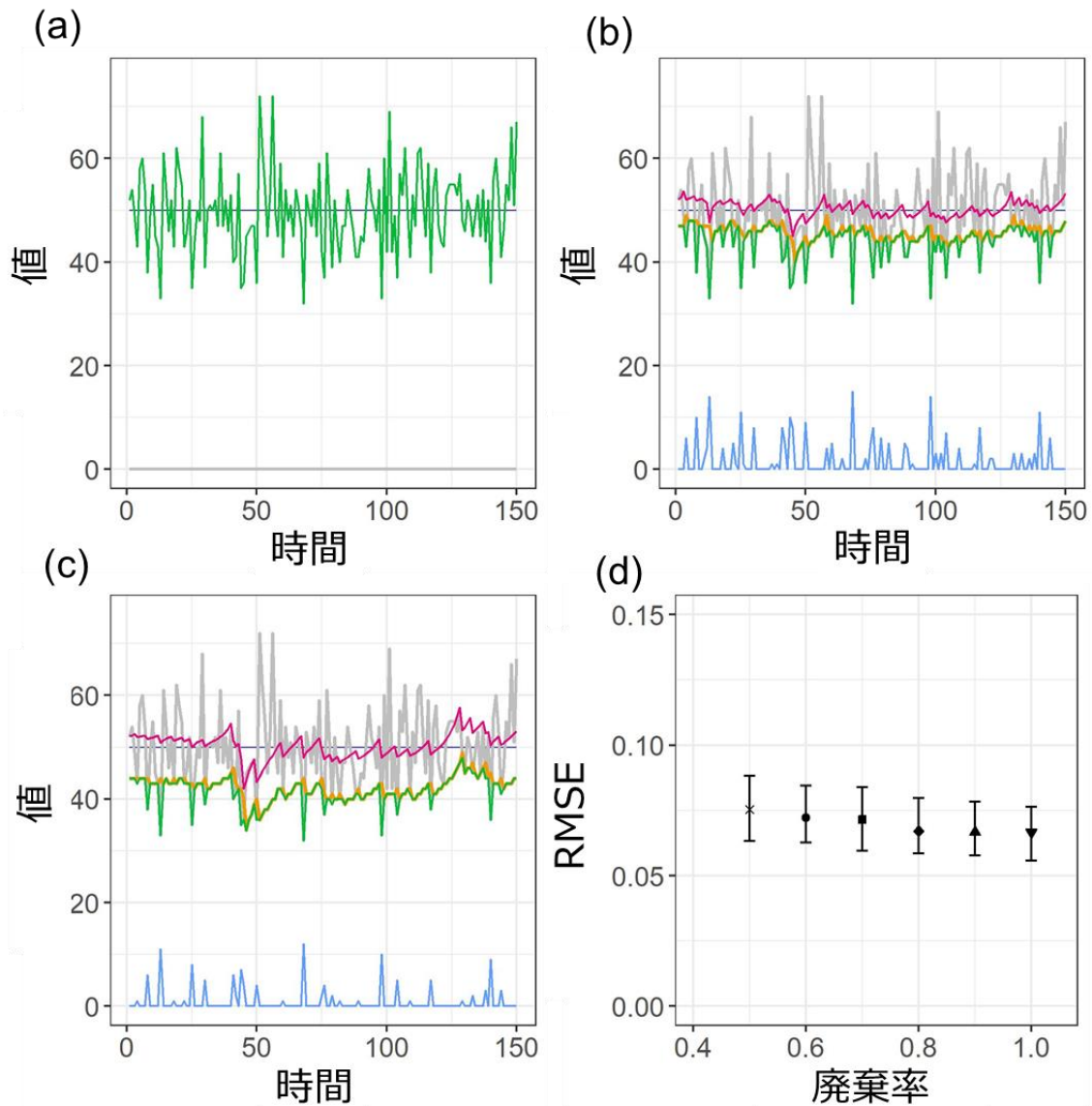


図 4.6 観測打ち切りがある定常時系列でのパラメータ推定の検証。(a) 平均 50、テイラースケーリングの比例定数 $\gamma=0.1$ 、式4.5の確率分布を仮定して生成した乱数時系列。緑線は需要、灰色線は需要平均。(b) 提案手法で狙い廃棄率 1.0 として得た販売時系列。灰色線は潜在需要、緑線は観測需要、橙線は仕入数、青線は廃棄数、赤線は推定需要平均。(c) 提案手法で狙い廃棄率 0.5 として得た販売時系列。(d) 狙い廃棄率毎の真の需要平均と推定需要平均の RMSE。黒点は RMSE の中央値、エラーバーは四分位範囲。

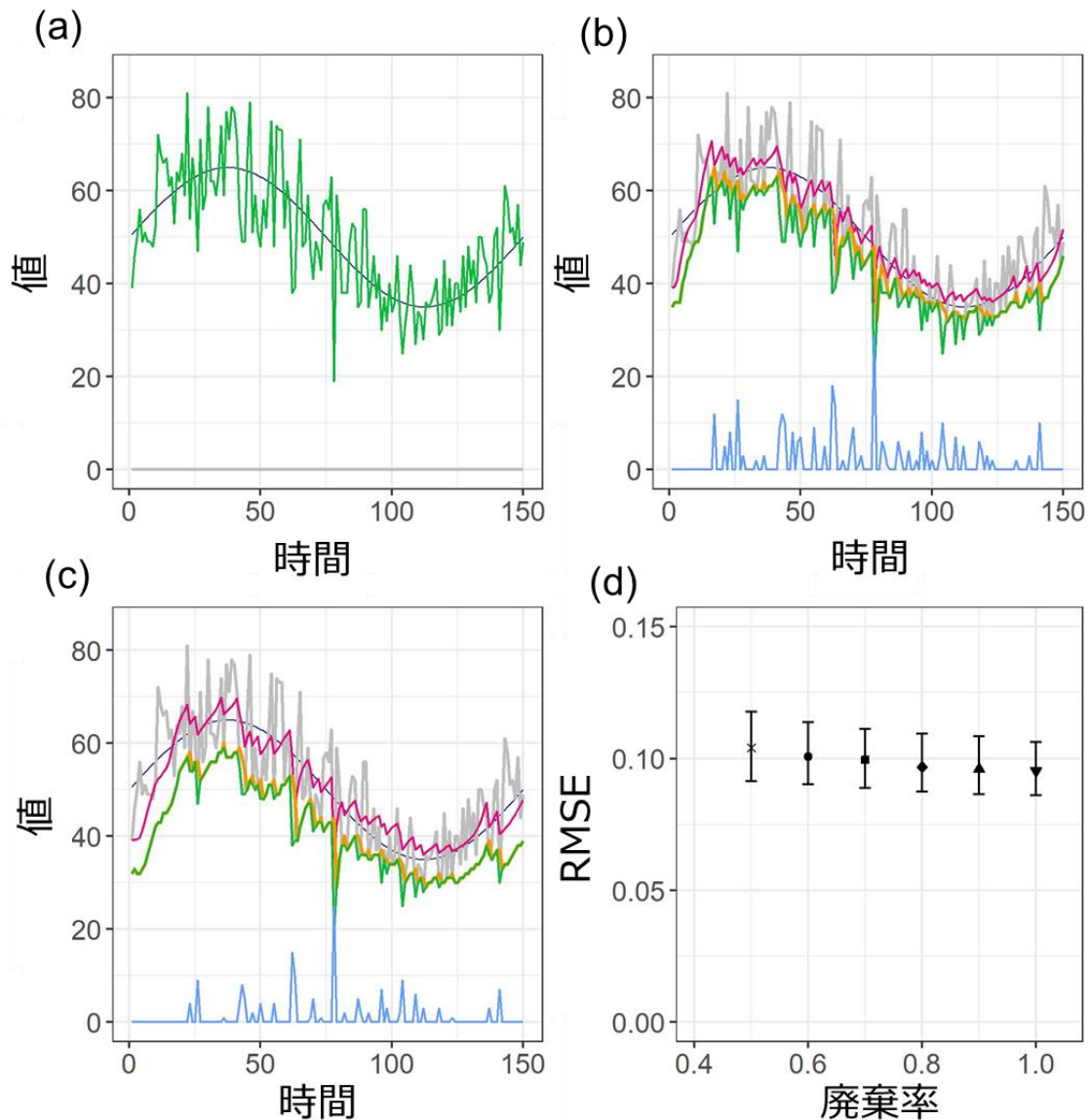


図 4.7 観測打ち切りがある非定常時系列でのパラメータ推定の検証。(a) 需要平均をサインカーブ（平均 50、振幅 30、周期 150）とし、テイラー・スケージングの比例定数 $\gamma=0.1$ 、式 4.5 の確率分布を仮定して得た乱数時系列。緑線は観測需要、灰色線は需要平均。(b) 提案手法で狙い廃棄率 1.0 として得た販売時系列。灰色線は潜在需要、緑線は観測需要、橙線は仕入数、青線は廃棄数、赤線は推定需要平均。(c) 提案手法で狙い廃棄率 0.5 として得た販売時系列。(d) 狙い廃棄率毎の真の需要平均と推定需要平均との RMSE。黒点は RMSE の中央値、エラーバーは四分位範囲。

表 4.1 検証に用いた店舗で調理した食品の POS データ

Product	Number of Shops	Total Disposal	Total Sales	Price (yen)
frankfurter sausages A	11	6030	22,200	150
frankfurter sausages B	8	2670	4291	150
frankfurter sausages C	2	635	1970	150
french fries A	22	8532	21,653	155
french fries B	25	11,939	30,350	155
french fries C	2	921	2615	165
fried chicken A	72	40,140	171,193	165
fried chicken B	33	17,098	152,790	105
fried chicken C	10	4464	14,869	105
grilled chicken A	3	1164	3075	105
grilled chicken B	12	4253	8651	105
grilled chicken C	13	4873	19,272	105
potato croquette	1	452	759	80
skewered beef	1	297	331	120
(Total number)	215	103,468	454,019	-

POS データによる検証

手法検証で用いる POS データは、153 日の販売日の内、75%以上の日で廃棄がなされた食品のものに限定した。その理由は 2 つある。一つは本研究の目的が廃棄低減であり、実際に発生した沢山の廃棄を低減できるのかを検証するためである。もう一つは、手法検証の技術的理由のためである。売り切れた日が少ない販売時系列は、販売数が真の需要を反映している。提案手法の需要推定結果を検証するためには、真の需要が既知である必要があるため、上述のような売り切れ日が少ない POS データを検証に用いることとした。選別された POS データは主に店舗で調理した食品、例えばフライドチキンやポテトなどのものであった。本章の手法検証で用いる POS データ 215 個を表 4.1 にまとめた。なお選別に際し、販売記録に欠損があるデータと、販売促進記録があって特異値を含むデータは除外した。

表 4.1 の POS データを用いて手法の効果検証を実施した。図 4.8(a),(b)はフライドチキン(Fried chicken A)の 2 つの店舗での販売時系列の例である。緑線は販売数を示している。図 4.8(a),(b)ともに上昇下降の非定常な需要変動がみられる。橙線は仕入数、青線は廃棄数である。図 4.8(a)では 153 日中 78%の日で廃棄が発生しており、327 個のフライドチキンが廃棄されている。153 日間で 1656 個のフライドチキンが 165 円で販売されており、売上は 273,240 円であった。仕入れた 1983 個のフライドチキンのコストは原価率を 0.7

とすると 229,037 円であるため、利益は 44,203 円と見積もることが出来る。図 4.8(b)では、153 日中の 76%の日で廃棄があり計 438 個のフライドチキンが廃棄されている。利益は 153 日間で 3387 個が 165 円で販売されたため 117,068 円である。提案手法によって、これらの数多くの廃棄を低減しながら、利益を確保できるのかを検証する。

図 4.8(c),(d)は、図 4.8(a),(b)と同じ POS データの販売数を真の需要とみなして、図 4.5 の手順により、各日の観測販売数と廃棄数から平均需要を推定し、翌日の仕入数を決定することを繰り返して得た販売時系列シミュレーション結果である。図 4.8(c),(d)の灰色線は、それぞれ図 4.8(a),(b)の販売数と同じものであり潜在需要を表している。緑線は観測された販売数、橙線は仕入数、青線は廃棄数、赤線は推定された需要平均である。図 4.8(c)では 153 日間での廃棄数は 69 個、利益は 54,599 円であった。図 4.8(a)の店舗データに対し、廃棄は 21.1% (69/327)に低減し、利益は 124% (54,599/44,204)に向上した。図 4.8(d)では、図 4.8(b)の店舗データに対し、廃棄は 43.9% (192/438)、利益は 93.7% (109,692/117,068)となった。

提案手法を表 4.1 の全ての POS データ 215 個に適用し、それぞれ廃棄と利益を見積もって図 4.9 にまとめた。図 4.9 の横軸は、店舗の販売時系列の廃棄を、提案手法で得た販売時系列の廃棄で割ったものである。横軸の値が灰色点線で示した 1 より大きい時、店舗の廃棄より提案手法の廃棄の方が小さい。縦軸は店舗の販売時系列の利益を、提案手法で得た販売時系列の利益で割ったものである。青点は、提案手法の利益の方が店舗の利益より大きい場合、赤点は店舗の利益の方が提案手法の利益より大きい場合を示す。縦軸の利益が負のものは、店舗の利益が負であった場合に対応している。なお、黒枠で囲った点は、青点が図 4.8(c)、赤点が図 4.8(d)の結果に対応している。提案手法により 215 個の販売時系列の 83.7% (180/215)で利益改善がみられることが分かった。215 個の販売時系列の廃棄と利益の結果を集計すると、店舗の廃棄数 (103,468) は提案手法により 23.2% (23,967)に低減し、店舗の利益 (8,255,002 円) は提案手法により 140%(11,558,387 円)に増加した。

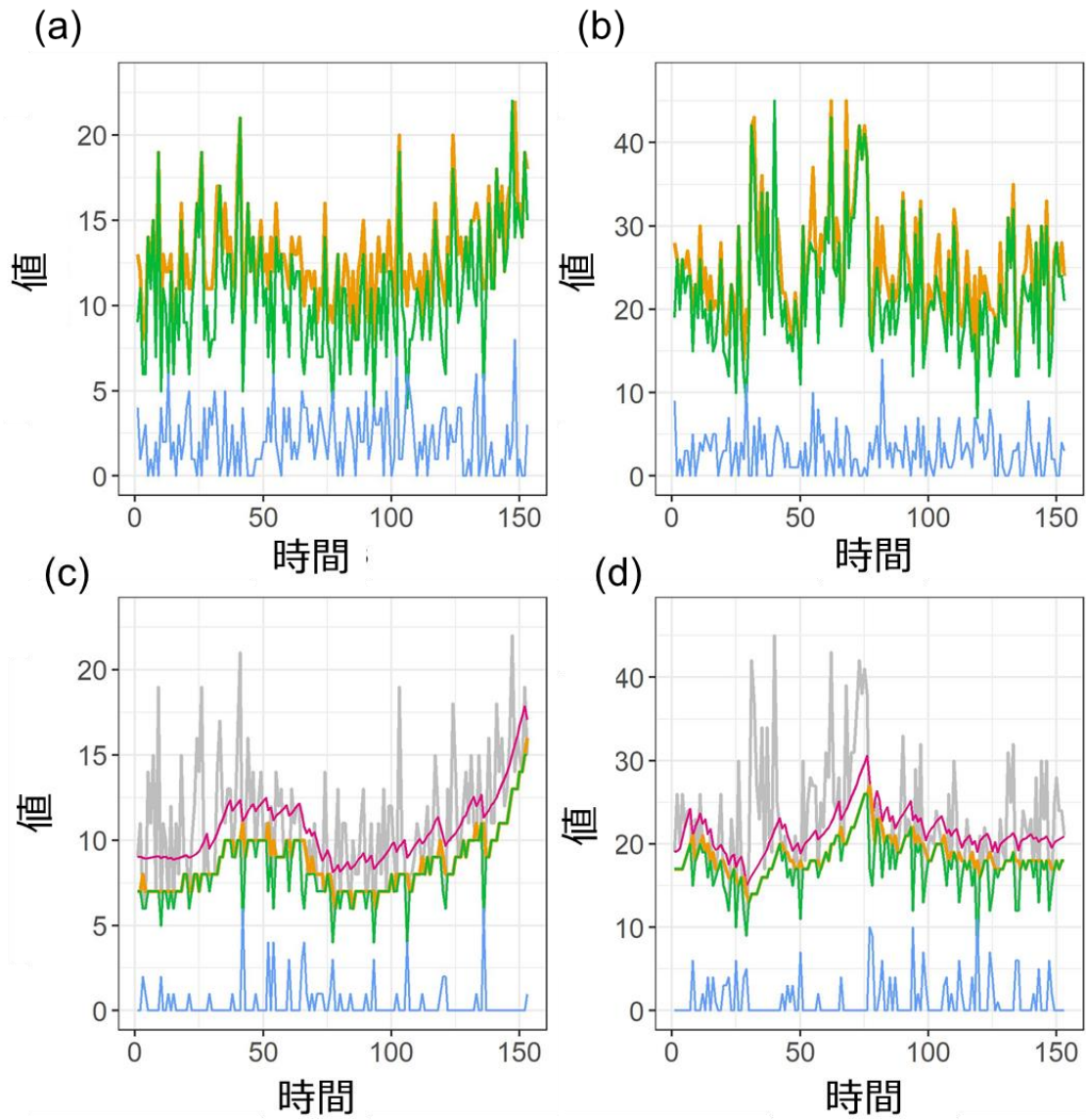


図 4.8 提案手法による販売シミュレーション結果。(a) ある店舗のフライドチキンの販売時系列。(b) ある別の店舗のフライドチキンの販売時系列。(c) (a)の販売時系列を潜在需要と見做して提案手法で販売時系列シミュレーションした結果。(d) (c)の販売時系列を潜在需要と見做して提案手法で販売時系列シミュレーションした結果。(a)から(d)の緑線は販売数、橙線は仕入数、青線は廃棄数、灰色線は潜在需要、赤線は提案手法で見積もった需要平均。

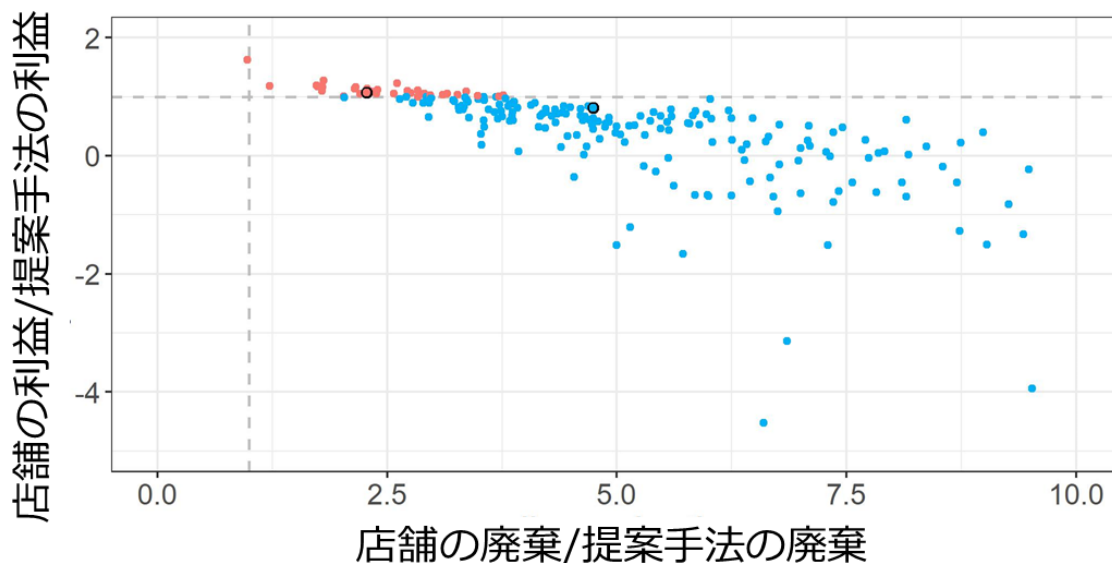


図 4.9 表 4.1 の POS 時系列 215 個による効果検証。店舗の利益より提案手法の利益が高かったものを青点、低かったものを赤点で表示。黒枠で囲った青点は図 4.8(c)、赤点は図 4.8(d)の結果に対応。

更なる廃棄低減の効果検証

ここまでの評価では、仕入数の決定に新聞売り子問題の式4.4を使っている。一方、提案手法では、式4.4の時の廃棄数に対して所望の廃棄率にできる仕入数を、その廃棄低減コストを見積もりながら決定する式4.9、式4.10を提供している。これらの式による更なる廃棄低減効果を検証する。表 4.1 の 215 個の POS データに対し、式4.9で日々の仕入数 $s(\alpha)$ を決定する販売シミュレーションを行った。狙い廃棄率 α は 0.5 から 1 まで 0.1 刻みで変化させた。図 4.10(a)は、狙い廃棄率 α 毎に得られた廃棄と利益の関係である。狙い廃棄率 α が 1、つまり新聞売り子問題の式4.4の仕入の時の利益と廃棄を 1 とし、横軸に廃棄比率、縦軸に利益比率を表示している。黒点は中央値、エラーバーは四分位範囲である。青の点線は式4.9、式4.10による理論線である。概ね理論線に従った廃棄と利益の関係が得られている。27%の廃棄低減は 1%の利益ロスで実現しており、45%の廃棄低減は 3.3%の利益ロスで実現している。

図 4.10(a)では、狙い廃棄率と実際に得られた廃棄率に若干の乖離がみられる。例えば狙い廃棄率 0.7 の時に得られた廃棄率は 0.73、狙い廃棄率 0.5 の時の得られた廃棄率は 0.55

である。この乖離は狙い廃棄率と得られた廃棄率の関係を回帰分析することで補正することが出来る。図4.10(b)はその回帰分析の結果である。回帰式は狙い廃棄率を $g(\alpha)$ 、得られた廃棄率を α とした時、 $g(\alpha) = 1.09\alpha - 0.09$ であった。この回帰式を用いて、廃棄率 α を得るための狙い廃棄率 $g(\alpha)$ を求めることが出来る。

図4.10(a)中の赤点は狙い廃棄率0.5から1に対して、補正した狙い廃棄率 $g(\alpha)$ を用いて得た仕入数を使った時の廃棄率と利益率の結果である。狙い廃棄率に相当した廃棄率が得られていることを確認できる。

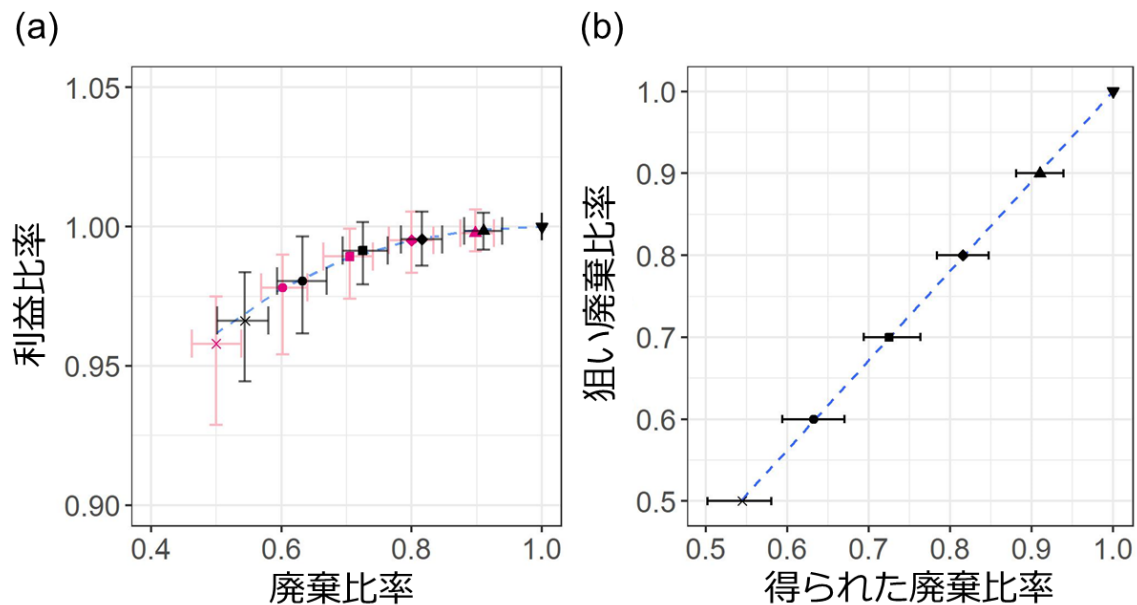


図 4.10 廃棄コストを見積もりながらの廃棄低減の検証結果。(a) 0.5 から 1 の狙い廃棄率に対して 215 個の POS データによる販売シミュレーションで得られた廃棄と利益。狙い廃棄率 1.0 の時の廃棄と利益で規格化して表示している。黒点は中央値、エラーバーは四分位範囲。赤点は(b)で決めた補正後の狙い廃棄率で得た廃棄率と利益率の中央値、赤エラーバーはその四分位範囲。青点線は式 4.9、式4.10による理論線。(b) 狙い廃棄率 $g(\alpha)$ と得られた廃棄率 α による回帰分析結果。青点線は回帰線 $g(\alpha) = 1.09\alpha - 0.09$ 。

原価率依存性

ここまでの検討では、仕入数や利益を求める際の原価率を 0.7 と仮定してきた。一方、実際には商品毎に原価率は異なる。それに加えて、様々な食品関連の業種への本手法の展開を考えると、様々な原価率に対して本手法の効果を見積もっておくことが重要である。表 4.1 の POS データ 215 個に対し、原価率を 0.1 から 0.9 まで 0.1 刻みで仮定した時の販売シミュレーションを行い、利益と廃棄を見積もった。仕入数は式 4.4 の新聞売り子問題の公式で決定した。図 4.11(a) は各原価率に対する 215 時系列の利益の集計結果である。緑線は店舗の利益、赤線は提案手法で得られた利益である。原価率 0.55 以上で提案手法の利益が店舗の利益を上回っている。図 4.11(b) は各原価率に対する 215 時系列の廃棄の集計結果である。原価率 0.25 以上で、緑線で表した店舗の廃棄より、赤線で表した提案手法で得られた廃棄が小さくなっている。原価率 0.55 では提案手法により、店舗と同等な利益を得ながら、店舗の廃棄を 40.6% に低減できている。

図 4.11(a) は、最大利益を与える新聞売り子問題の公式 4.4 を使用しているにも関わらず、原価率 0.55 以下で提案手法による利益が店舗の利益を下回った。その原因の一つとして、店舗の販売時系列では 153 日の内の最大で 25% の日で売り切れが発生しているため、提案手法による利益向上効果が過小評価されていることが挙げられる。この効果については後に図 4.13 で検討する。もう一つの原因として、提案モデルで想定されていない要因が店舗の販売データに含まれていることが挙げられる。図 4.12(a),(b) はその考察のための販売時系列である。緑線は販売数、赤線は原価率 0.1 とした時の推定需要平均である。需要分布の揺らぎに対して、大きな販売ピークが発生していることが分かる。式 4.5 の需要分布 $m(k|\lambda)$ に対し日々の販売数が発生する確率を求めると、図 4.12(a),(b) で赤点で示した販売数は CDF で 0.95 より大きな値となる。このような CDF 値 0.95 を超える販売ピークが、理論で想定されるよりも多く発生している。図 4.12(c) は 215 販売時系列の各日の販売数の CDF 値のヒストグラムである。確率分布がデータと整合しているとき CDF の各値の発生頻度は一様になるのに対し、CDF 値が 0.95 以上の頻度は、他の値の頻度の 2 倍程度となっている。図 4.12(d) は 215 個のそれぞれの時系列で CDF 値 0.95 以上の観測値の個数を集計し、ヒストグラム表示したものである。販売日数 153 日の中で 15 日以上 CDF 値 0.95 以上の観測値が得られる確率は 1.0% である。しかし、実際には 84.2% の販売時系列で CDF 値 0.95 以上の観測値が 15 日以上発生している。100 σ 以上離れた販売数が観測されている時系列も一つ存在した。これらの CDF 値 0.95 以上の販売ピークは、イベント向けのまとめ

発注など、例外的な状況で発生していると考えられる。需要分布の揺らぎに対して低い確率で発生する需要に対し、売り切れなく販売できていることから、事前の予約などの何らかの事前情報が店舗にもたらされていたと考えられる。

提案手法の潜在的な可能性を評価するため、CDF 値 0.95 以上の観測値が 15 日以上発生している販売時系列、及び 100 σ 以上離れた販売数が観測された販売時系列を除外し、得られた 33 個の販売時系列による販売シミュレーションを行って廃棄と利益を見積もった。図 4.13(a),(b)はその廃棄と利益の見積もり結果である。緑線は店舗のデータ、赤線は提案手法の結果を示している。青線は、POS データの売り切れ日での販売数補正を実施した場合の、提案手法の結果である。売り切れ日での需要の値は、仕入数以上の販売数について需要分布との積和をとり、その値を仕入数以上の販売数の需要分布の和で割ることで求めた。この需要の値までは売れると見做し、提案手法で決めた仕入数に対する販売数を求めた。図 4.13(a)の青線より、売り切れ補正を実施した時の提案手法による利益は、原価率が低い時に店舗の利益に漸近していくことが分かる。また、図 4.13(b)より、売り切れ補正を実施した時の提案手法による廃棄は、売り切れ補正をしない時の提案手法による廃棄よりも小さくなる事が分かる。

図 4.13(c)は上述の 33 個の販売時系列でみた観測販売数の CDF 値のヒストグラムである。CDF 値 0.95 以上の観測値は抑制されていることが分かる。データ数が少ないため、ヒストグラムは一様になっていない。図 4.13(d)は 33 個の販売時系列で日々の需要平均 λ を求め、式 4.5 の需要分布 $m(k|\lambda)$ に従った乱数を生成したシミュレーション結果である。図 4.13(c)の実データの CDF 値の分布と図 4.13(d)の理論的な CDF 値の分布の整合性を Kolmogorov-Smirnov 検定により確認した。理論的な CDF 値は乱数シードを変更して 200 セット準備して実データの CDF 値と比較した結果、p 値は平均 0.07 であった。この結果は実データと理論的な CDF の分布の差異は有意確率 0.05 に対し有意であるといえないことを示している。

図 4.13(a)は、提案手法による利益と廃棄が原価率 0.2 で店舗のそれらと等しくなり、原価率 0.4 では提案手法と店舗の利益は同等で、廃棄は提案手法により店舗の 59.7% (8247/13,822)になることを示している。また原価率 0.7 では提案手法による利益は店舗の 208.0% (1,260,691 / 606,070 円)、廃棄は 22.4% (3097 / 13,822)になっている。

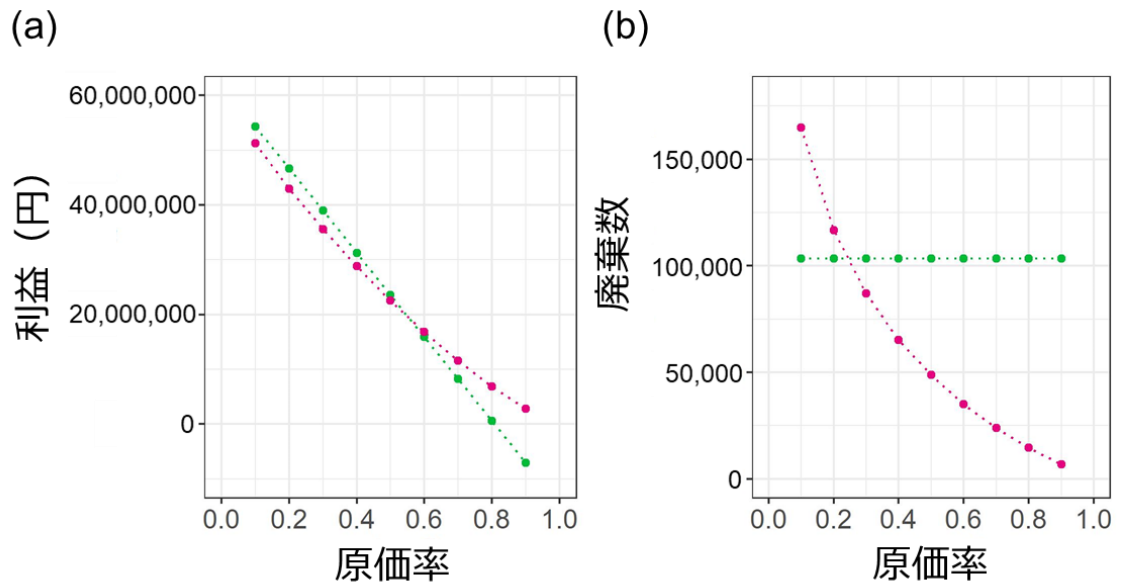


図 4.11 各原価率に対し 215 個の POS データで販売シミュレーションを行った際の利益と廃棄の集計結果。(a) 各原価率に対して得られた利益。(b) 各原価率に対して得られた廃棄。緑線は店舗の利益と廃棄、赤線は提案手法により得られた利益と廃棄。

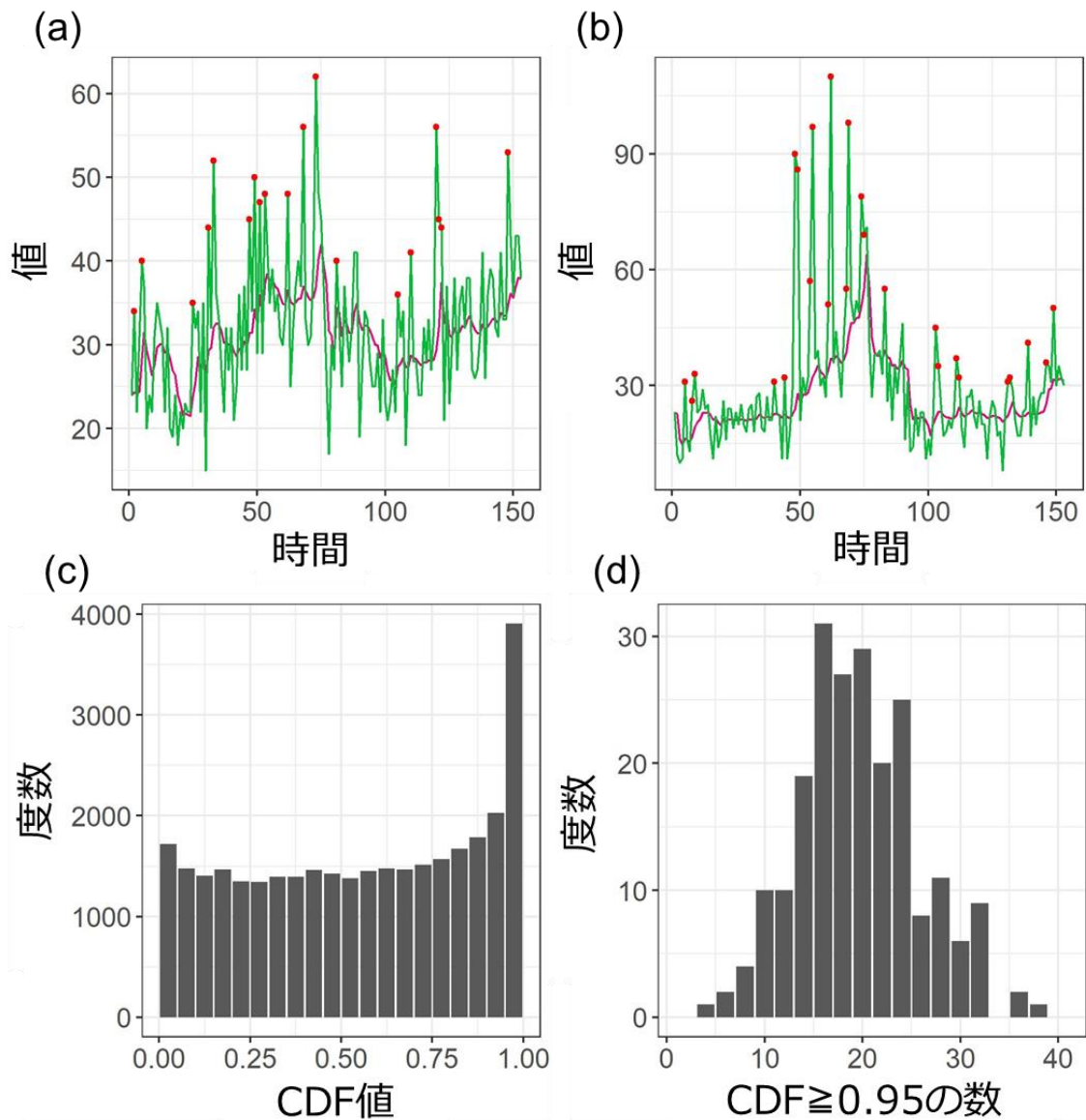


図 4.12 販売時系列の中の急峻なピーク。(a) CDF 値 0.95 以上の販売が 19 個観測された販売時系列。(b) CDF 値 0.95 以上の販売が 25 個観測された販売時系列。(c) 215 個の POS データの各日の販売数の CDF 値のヒストグラム。(d) 215 個の POS データのそれぞれで集計した CDF 値 0.95 以上の観測値の個数のヒストグラム。(a),(b)の緑線は販売数、赤線は原価率 0.1 とした時の推定需要平均。赤点は CDF 値が 0.95 以上の観測販売数。

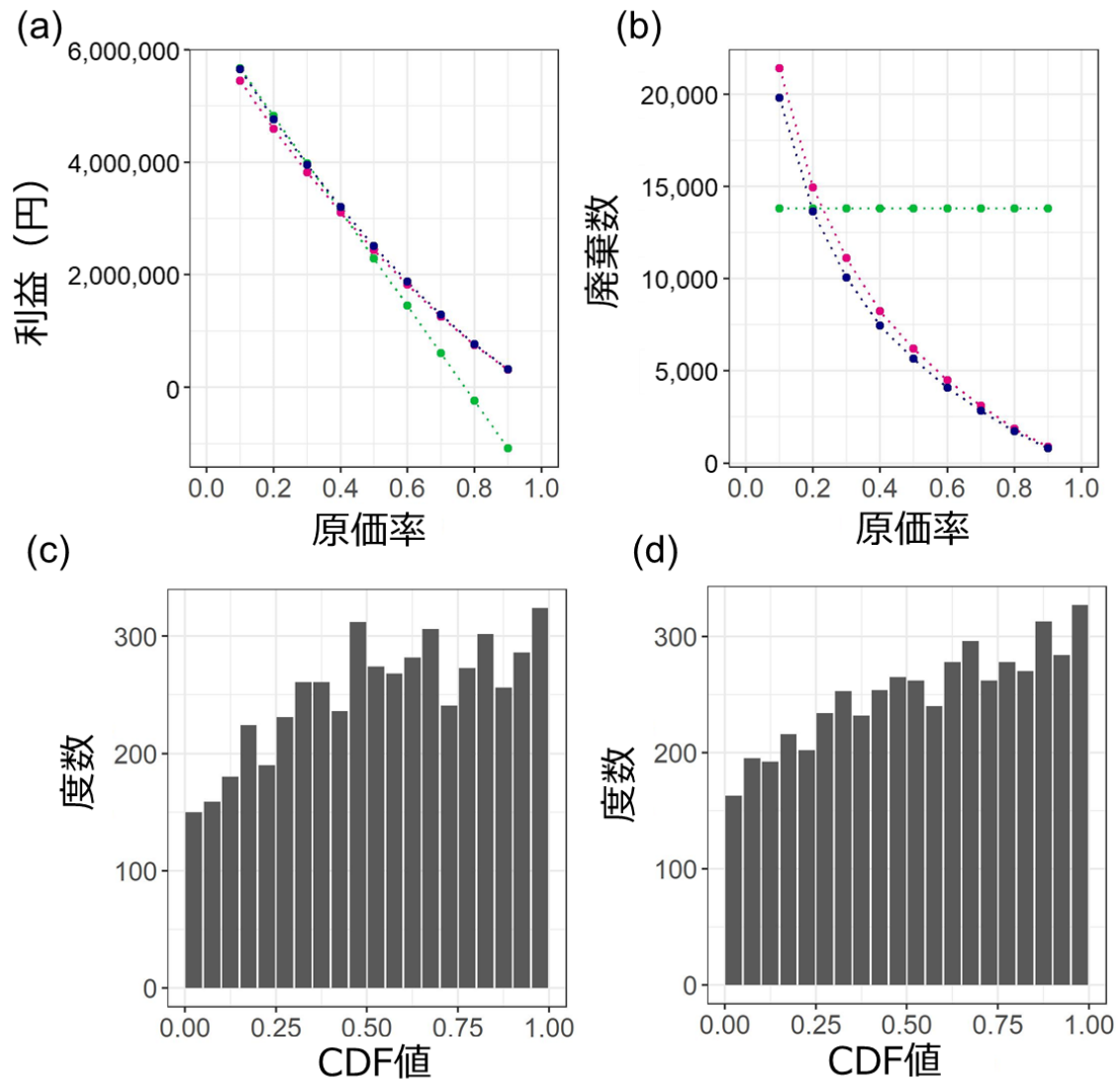


図 4.13 CDF 値 0.95 以上の観測値が 15 日以上発生、または 100σ 以上離れた観測値を含む POS データを除外した、33 個の POS データによる販売シミュレーションで見積もった廃棄と利益。(a) 各原価率に対して得られた利益。(b) 各原価率に対して得られた廃棄。緑線は店舗の利益と廃棄、赤線は提案手法により得られた利益と廃棄、青線は売り切れ補正を実施した時の利益と廃棄。(c) 33 個の POS データでみた観測販売数の CDF 値のヒストグラム。(d) 33 個の POS データで日々の需要平均 λ を求め、式 4.5 の需要分布 $m(k|\lambda)$ の乱数で観測値を生成して得た理論的な CDF 値のヒストグラム。

従来手法に対する優位性

ここで提案手法と従来手法の比較を行う。第一章で述べたように打ち切りがある販売データから需要推定し仕入数を決定する方法としてBDP [59, 60, 61]やそれと関連した近似的手法 [64, 65, 66, 67]、ノンパラメトリック手法 [71, 72, 73, 132]が提案されている。一方、これらの手法は非定常な場合には対応していない [75]。非定常な需要にはARIMA [36, 37, 38, 39]、NN [40]、EWMAなどが対応できる。一方で、これらの非定常時系列分析手法は打ち切りデータに対応していない。

ここでは非定常で打ち切りがある販売時系列に対してARIMA、NN、EWMAを適用した場合の結果を示す。ARIMAとNNによる分析は、統計分析ソフトウェアRのforecastとnnetパッケージをそれぞれ用いて実施した。EWMAは、 D_t を時間 t での推定需要、 Y_{t-1} を時間 $t-1$ での観測販売数、 α を平滑化係数として、 $D_t = \alpha Y_{t-1} + (1 - \alpha)D_{t-1}$ によって計算した [35]。テストデータは図4.8(a)の販売データとした。ARIMAとNNは学習を必要とするため、最初の90日間の販売データは学習に用いた。学習時には打ち切りはないものとして販売データを扱った。91日目以降は、得られた学習モデルを用いて販売データから需要推定を行った。日が進み、新たな販売データが得られる度にその販売データを追加して学習モデルを更新した。簡単のため、仕入数は推定された需要と同じ値に設定した。これは原価率が0.5で、需要分布が対称であるときの最適仕入数と同等である。学習されたARIMAモデルは赤池情報量基準(AIC)より $(p, d, q) = (0, 1, 1)$ となった。 p は自己回帰(Autoregressive: AR)過程の次数、 d は和分(Integrated)過程の次数、 q は移動平均(Moving Average)過程の次数である [36]。NNに関しては、隠れ層の中のニューロン数は10とした。ニューロン数が多いとオーバーフィッティングと不安定な推定に繋がり、ニューロン数が少ないとアンダーフィッティングとなる。ニューロン数10は、テストデータに対するフィッティングと推定の安定性を事前に調査して選択した。EWMAは学習を必要としない。1日目の需要推定値 D_1 は販売数観測値 Y_1 とし、 α は事前の条件検討から0.12として2日目の需要から推定を行った。

図4.14(a),(b),(c)はそれぞれARIMA、NN、EWMAによる需要推定結果である。灰色線は潜在的な需要、青線は打ち切りがないと仮定して潜在需要をそのまま観測値とした時の推定需要である。いずれの手法でも、打ち切りがない場合の推定需要は概ね時系列の非定常な変化に追従している。緑線は仕入数による観測打ち切りがあるとした時の観測値である。赤線は打ち切りのある観測値からの推定需要である。赤線は時系列の推移に追従できてい

ないことが分かる。ARIMA、NN、EWMA は打ち切られた観測値をもとに需要を低く見積もってしまう。これら従来手法は打ち切られたデータからの推定に対応しないため、何らかの観測打ち切りに対する補正が必要になる。例えば、売り切れ時には勘と経験により本来売れたであろう量を推測すること、販売データの何らかのパターンを補正に使うことなどである。一方、提案手法では打ち切りの補正を最尤推定法により統計的に行う。勘と経験の適用やデータの潜在的なパターンの想定は不要である。

提案手法において、テイラースケーリングを考慮した時としない時の比較も実施した。図4.14(d)はその結果である。図4.14(d)の時系列の平均値は概ね50程度と大きく($\lambda > 20$)、テイラースケーリングの効果が想定される。青線は粒子フィルタの尤度関数に式3.22を用いてテイラースケーリングを考慮した場合の推定結果である。赤線は粒子フィルタの尤度関数に式3.21を使いテイラースケーリングを考慮しない場合の推定結果である。テイラースケーリングを考慮しない場合には、考慮した場合より推定需要が小さく、利益は3.3%小さくなることを確認した。テイラースケーリングによる大きな揺らぎは、ポアソンモデルでは非定常変化とみなされる。観測値に打ち切りがある場合、高い側の分布の裾は見られず、低い側の分布の裾が見られるため、推定は低い側にずれることとなる。

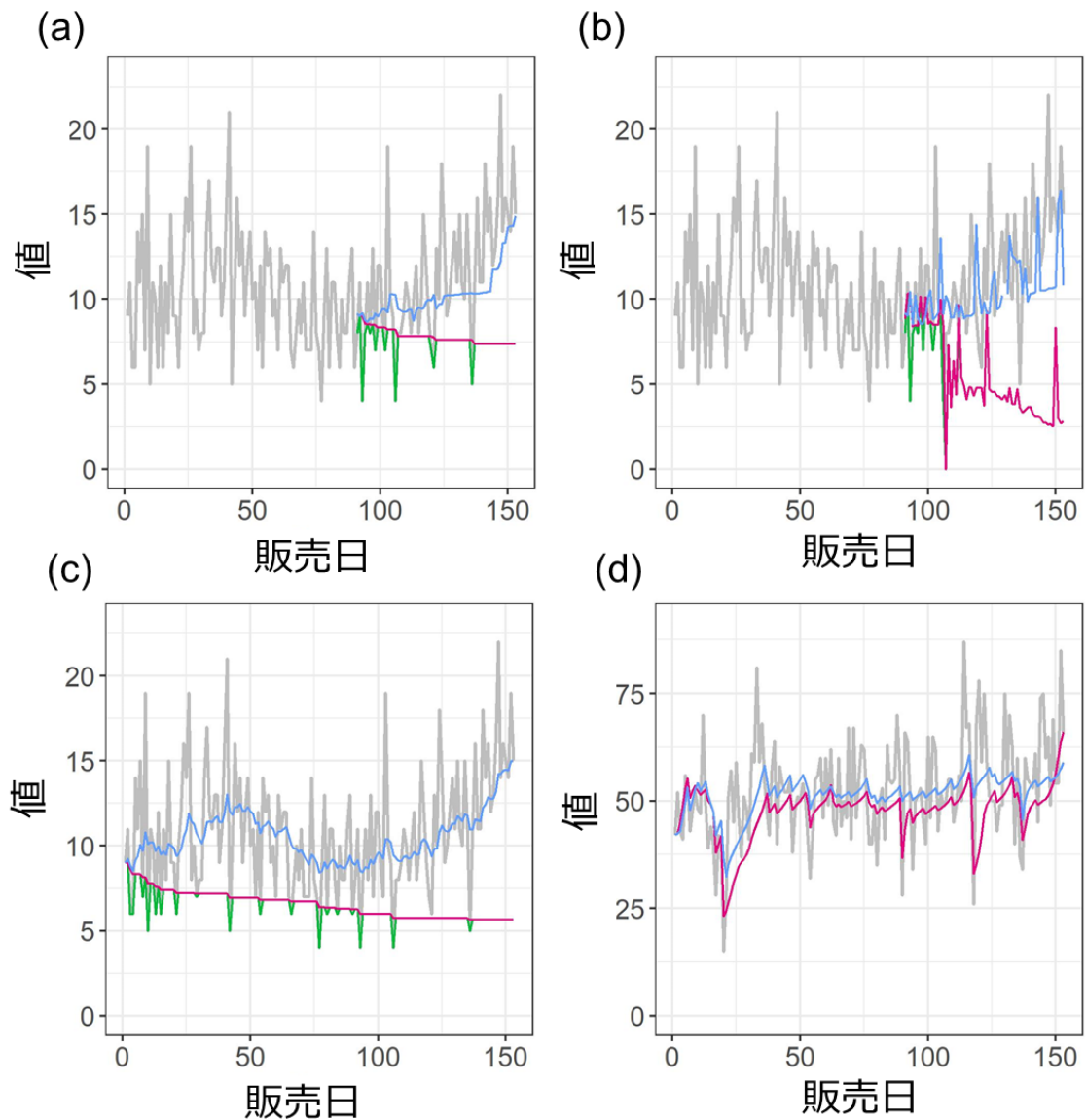


図 4.14 従来手法による需要推定結果。(a) ARIMA による推定結果。(b) NN による推定結果。(c) EWMA による推定結果。(d) 提案手法でテイラースケーリングを考慮した場合としない場合の推定結果。(a)から(c)について、灰色線は潜在的な需要、青線は打ち切りがないと仮定して潜在需要をそのまま観測値とした時の推定需要、緑線は仕入数で打ち切られた観測値、赤線は打ち切られた観測値からの需要推定結果。(d)について灰色線は潜在需要、青線はテイラースケーリングを考慮した時の推定結果、赤線はテイラースケーリングを考慮しない場合の推定結果。

4.4 多店舗多商品の集計販売時系列での効果検証

本研究では、所望の廃棄率にできる仕入数を、廃棄低減コストを見積もりながら決定するための式4.9、式4.10を提案した。ここでは、これらの式を詳細に検討する。図4.15はテイラースケーリングの比例定数 γ と需要平均 λ を変化させた時の廃棄率と利益率の関係である。狙い廃棄率1.0の時の利益と廃棄を1に規格化している。青線は $\lambda = 10$ 、橙線は $\lambda = 3000$ 、 $\gamma = 0.3$ 、赤線は $\lambda = 3000$ 、 $\gamma = 0.12$ 、緑線は $\lambda = 3000$ 、 $\gamma = 0.05$ の時の結果である。青線の需要平均10の時はテイラースケーリングの効果が殆ど現れないが、赤線、橙線、緑線の需要平均3000の時はその効果が明瞭に表れる。比例定数 γ が小さくなるほど廃棄低減に必要な利益ロスが小さくなることが分かる。具体的には、廃棄半減に必要な利益ロスは、 $\gamma = 0.3$ の時には3.9%程度であるのに対し、 $\gamma = 0.12$ では1.2%程度、 $\gamma = 0.05$ では0.5%程度となっている。

比例定数 γ が小さくなることによる利益ロスの減少には、 γ が小さくなると需要分布の揺らぎが小さくなることと関係している。平均が10の時の標準偏差 σ は平均に対し31.6%、 $\gamma = 0.3$ の時は $\sigma = 30.1\%$ 、 $\gamma = 0.12$ の時は $\sigma = 12.1\%$ 、 $\gamma = 0.05$ の時は $\sigma = 5.3\%$ である。需要分布の揺らぎが小さいほど、廃棄を減らすための仕入減は少なくて済み、販売数並びに利益を高く維持できることとなる。

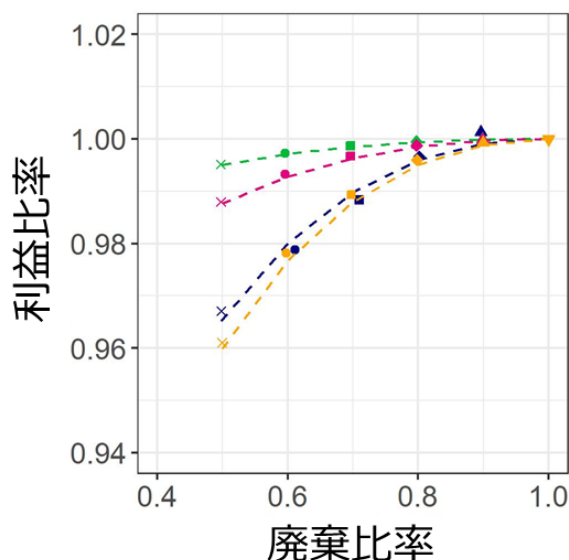


図 4.15 ティラースケーリングの比例定数 γ と需要平均 λ を変化させた時の狙い廃棄率と利益の関係。緑線は $\lambda = 3000$ 、 $\gamma = 0.05$ 、赤線は $\lambda = 3000$ 、 $\gamma = 0.12$ 、橙線は $\lambda = 3000$ 、 $\gamma = 0.3$ 、青線は $\lambda = 10$ 。

本研究で用いたコンビニエンスストアのPOSデータの販売数は、日々、数十個から数百個程度である。需要平均が大きい時の提案手法による廃棄低減効果を検証するため、表4.1のPOSデータを集計して、販売平均が大きい販売時系列を得た。集計に際しては、表4.1中のfrankfurterAといった同じ商品の複数店舗の販売データを日々合算、またfrankfurterAからCといった同じ商品種類の複数店舗の販売データを日々合算、また全販売データを日々合算することで、19個の販売時系列を得た。図4.16は19個の販売時系列の平均と標準偏差の関係である。実線は $\gamma = 0.12$ とした時のテイラースケーリングの関係 $\sigma = \sqrt{\lambda + (\gamma\lambda)^2}$ 、点線は $\sigma = \sqrt{\lambda}$ 、一点鎖線は $\sigma = \gamma\lambda$ である。集計した販売時系列の平均と標準偏差は概ねテイラースケーリングの関係に従っていることが分かる。

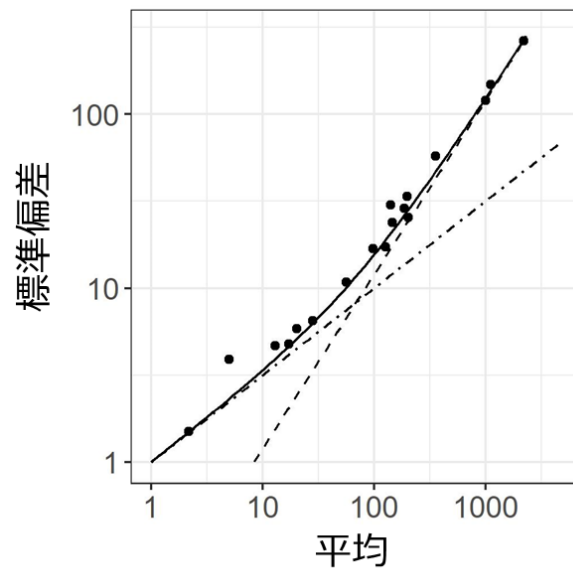


図 4.16 表 4.1 の POS データを集計して得た 19 個の販売時系列の平均と標準偏差の関係。実線は $\gamma = 0.12$ とした時のテイラースケーリングの関係 $\sigma = \sqrt{\lambda + (\gamma\lambda)^2}$ 、点線は $\sigma = \sqrt{\lambda}$ 、一点鎖線は $\sigma = \gamma\lambda$ 。

これら 19 個の販売時系列で提案手法による販売シミュレーションを行い、廃棄と利益を見積もった。図 4.17(a)は平均の異なる 19 個の販売時系列それぞれに対する廃棄の結果である。緑点は店舗の廃棄、赤点は狙い廃棄率を 1.0 とした時の提案手法による廃棄、青点は狙い廃棄率 0.5 (図 4.10(b)の回帰式により補正後の狙い廃棄率 0.455 を設定) の時の提案手法による廃棄である。図 4.17(b)は、狙い廃棄率 1.0 の赤点の廃棄に対し、狙い廃棄率 0.5 の青点の廃棄が概ね半分になっていることを確認するために、各時系列で青点の廃棄数を赤点の廃棄数で割った結果である。廃棄は中央値で 48.3%と概ね半分であった。

図 4.17 (c)は、19 個の販売時系列それぞれに対する利益の結果である。各時系列の狙い廃棄率 1.0 の赤点の利益に対し、狙い廃棄率 0.5 の青点の利益の比率を求めるため、各時系列で青点の利益を赤点の利益で割り、図 4.17(d)に示した。利益の中央値は 99.4%であった。データ数が 19 個と少ないため、おおよその傾向の確認に留める必要があるものの、概ね半分の廃棄が 1%前後の利益ロスで得られることが確認できた。

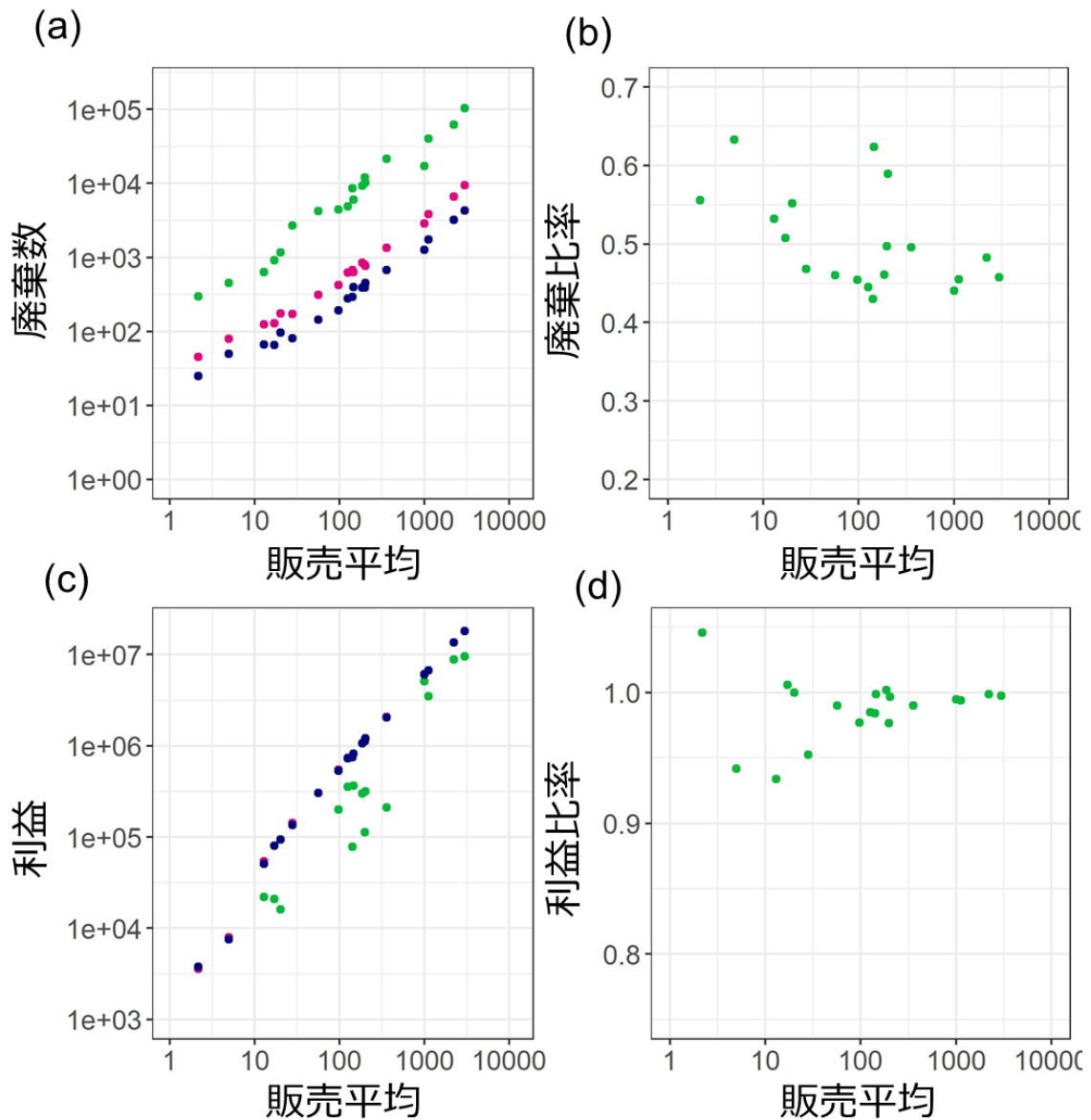


図 4.17 集計して得た 19 個の販売時系列による販売シミュレーション結果。(a) 各販売時系列の廃棄の結果。緑点は店舗データ、赤点は狙い廃棄率 1.0 の時の提案手法の結果、青点は狙い廃棄率 0.5 の時の提案手法の結果。(b) 提案手法で狙い廃棄率 0.5 の時の廃棄を、狙い廃棄率 1.0 の廃棄で割った結果。(c) 各販売時系列の利益の結果。緑点は店舗データ、赤点は狙い廃棄率 1.0 の時の提案手法の結果、青点は狙い廃棄率 0.5 の時の提案手法の結果。(d) 提案手法で狙い廃棄率 0.5 の時の利益を、狙い廃棄率 1.0 の利益で割った結果。

手法の実用性についての考察

小売店では、日々、沢山の商品の仕入数を決める必要がある。提案手法は、非定常な販売データから自動的に仕入数を決定することを可能にする。一定の在庫水準や、勘と経験による仕入れ [30]を抑制でき、廃棄低減や利益向上に寄与できると期待される。

小売店への新たな手法の導入に際しては、多角的な視点から導入の制約を考察することが重要であると指摘されている。それは、小売店の内的要因（インフラ、スペース等）、人的要因（店員の数、スキル、モチベーション等）、外的要因（顧客や供給者からの協力等）、法的要因（法律や条例等）である [103]。これらの内、内的要因、外的要因、法的要因に関しては、提案手法では問題にならないと考えられる。分析する販売データは店舗自身のものである。データ分析も大規模な計算環境を必要とせず、一般消費者向けの PC で実施できる。一方、人的要因については考慮が必要な点がある。提案手法の実装には、情報処理技術やデータ分析技術の知識を有する人材が必要となる。小規模な小売店においては特に、これらの知識を有する人がいないことを想定する必要がある。簡易的な操作で提案手法を使うことを可能とするソフトウェアや、オンラインサービスを提供することは現実的な解になると考えられる。その他要因として、売り切れは機会損失だけでなく、顧客の店に対する信頼、ロイヤリティを失う恐れがあるものである [30]という考え方も提案手法の導入の妨げになる。提案手法は過剰な在庫を抑制することで廃棄を低減するため、商品が売り切れとなる割合を増やすためである。廃棄低減を実現するための売り切れの考え方について議論が必要である。その他、一般に小売店経営者は、情報技術を知ることやツール導入に対して積極的ではないことも指摘されている [103]。提案手法の成功事例を積み重ね、利益向上と廃棄低減の利点を明示していく必要がある。

バイオマスや再生可能エネルギーなど多大な投資を必要とする環境技術の導入では、銀行やベンチャーキャピタルなどが重要な役割を果たしている [133]。提案手法のソフトウェアや PC の導入にはそのような多大な投資は必要とされないものの、小規模の小売店にとっては、経済面も提案手法の導入の障壁となりうる。環境技術導入のための小規模の融資や助成金の事例は少なく [134]、こういった小規模な経済支援も、提案手法を含む様々な環境技術の社会実装を促進すると考えられる。

4.5 まとめ

第3章で述べたテイラースケーリングに従う非定常ポアソン時系列のパラメータ推定手法を拡張して小売店での需要推定と仕入決定に適用可能とした。提案手法にテイラースケーリング則を考慮した打ち切りデータの最尤推定法を導入することで、売り切れを含んだ不完全な販売データからの需要推定を可能とした。推定需要から仕入決定するにあたって、機会損失と廃棄損失の最適化問題を解くための新聞売り子問題の公式を拡張し、所望の廃棄率を実現できる仕入数を、利益ロスを見積もりながら決定可能とした。75%以上の販売日で廃棄がなされた食品の販売時系列 215 事例において、平均的に廃棄数を 23%に減少させながら利益を 140%に向上させる効果を確認した。さらに、複数店舗、複数商品の POS データを集計した販売時系列を用意して、テイラースケーリングの効果が明瞭となる販売数平均が大きい場合の効果検証を行い、販売平均が大きくテイラースケーリング比例定数 γ が小さい時、例えば $\gamma=0.12$ の時には、最大利益を得られる仕入数の時の利益と廃棄に対し、1%程度の利益ロスで廃棄を半減できるという結果を得た。

第5章 販売データの統計的性質とその起源 の解明

第5章では、商品コード毎の販売店舗数の変動と分布に関する数理モデルの開発を行う。まず、POS データの分析から、全 326 店舗で扱われる全商品コードは一日当たりおよそ 8000 種程度であり、その内およそ 500 種が毎日入れ替わっていることを確認し、また、商品コード毎の販売店舗数は、定常的に対数一様分布に従うという経験則を示す。さらに、商品コード毎の販売店舗数の時系列の解析から、商品コード毎の販売店舗数の変動がランダム乗算過程で近似できることを示し、数値シミュレーションによって、一般にランダム乗算過程に従って有限な区間内を変動することで対数一様分布が実現することを明らかにする。これらの結果を踏まえ、商品の入れ替わりを考慮したランダム乗算過程に基づく数理モデルによって販売店舗数の変動と分布の特性を再現することができることを確認する。なお、本章の内容は論文 [135]に基づく。

5.1 ベキ分布の従来研究

本章で扱う対数一様分布はベキ分布と関連している。本節では、ベキ分布の従来研究を概観する。ベキ分布とその生成メカニズムは100年以上も前から研究されてきている。1916年にスモルコフスキーはコロイド粒子の質量の分布において見出されたベキ分布の起源を考察した [136]。その後の数々の研究 [137, 138, 139, 140, 141, 142]によりこの分布は、小さな質量の粒子の継続的な注入がある条件下での、非可逆の凝集過程によって生み出されることが確認された。ベキ分布は相転移の臨界点、例えばパーコレーションにおけるクラスターサイズ分布においてもみられる。自己組織化臨界現象 [143]は、そのような臨界点に向かうパラメータ調整が自動的になされることを示すものである。雪崩 [143]、太陽フレア [144]、地震 [145]において確認されるベキ分布は、自己組織化臨界現象の具体例である。物理学者の研究対象は、経済や社会現象にまで拡大しており、金融市場の価格変化 [146]、企業の売り上げ分布 [147]、また企業間の取引関係 [13]などにおいてもベキ分布が見出されている。

数理的な手法によるベキ分布の研究例も多くみられる。1925年に Yule は生物学的分類の中で見出されたベキ分布を説明するために、優先選択を提唱した [148]。優先選択は、後に単語の使用頻度 [149]、また複雑ネットワークのリンク数におけるベキ分布 [150]の説明に適用された。1930年代に Levy は中心極限定理を一般化し、独立した確率変数 x_i の和 S_n について、 x_i の分散が発散し $|S_n|$ が十分に大きい時には $|S_n|$ がベキ分布に従うことを安定分布の理論で示した [151]。ランダム乗算過程、または係数を確率分布に従って変化させたランジュバン方程式もまた、ベキ分布を生成することが知られている [152, 153]。

ベキ分布において数々の研究がなされてきたものの、これまで注目されてこなかったベキ分布の特殊形がある。それは、ベキ指数が0のベキ分布である。その確率分布は変数の対数をとると一様となるため、対数一様分布と呼ばれる。対数一様分布は、統計的な考察やコンピュータシミュレーションでの仮定として使われるだけのものではなく、自然、社会現象の実際の分布として確認されている。具体的には、宇宙物理学において、クラス1の原子星の連星間距離は対数一様分布に従うと見做されている [154]。地球科学の分野では、カナダの始生代断層帯に沿って曲線的に測られた金の鉱床間距離の分布は対数一様分布となっていることが確認されている [155]。人の行動についても、計算機ユーザーが見積もるジョブの実行時間の分布が対数一様分布であったという報告がある [156]。このような対数一様分布の実例の報告は、ベキ指数が0でないベキ分布と比較して少なく、対数一様

分布の起源も未だ明らかとなっていない。本章では、POS データの詳細な分析から、商品コード毎の販売店舗数が定常的に対数一様分布に従うことを報告するとともに、数理モデルを構築して対数一様分布の起源を解明する。

5.2 商品の入れ替わりの激しさ

POS データの分析により、商品の活発な入れ替わりの様子が明らかとなる。図 5.1(a)は全 326 店舗における商品コード数の推移である。青線は日々の全商品コード数であり 8000 個程度で揺らいでいる。緑線は前日になかった商品コードが増えた数（参入と呼ぶ）を日々集計したものである。赤線は前日あった商品コードが減った数（退出と呼ぶ）を日々集計したものである。日々 500 個程度、定常的に商品コードの参入、退出があることがわかる。なお、各商品コードの参入と退出は度々起こるため、全 153 日での重複なしの商品コードの数は 21,037 である。

商品コード毎の販売店舗数の推移からも商品の入れ替わりの激しさが見て取れる。図 5.1(b)は商品コード毎の販売店舗数の推移の例である。赤線のように販売店舗数が定常的に推移する商品コードがある一方で、青線のように販売のない状態からある日に全店舗で販売がなされ、ある日に全店舗で販売がなくなるものや、緑線のように山なりに販売店舗数が推移する商品コードが存在している。

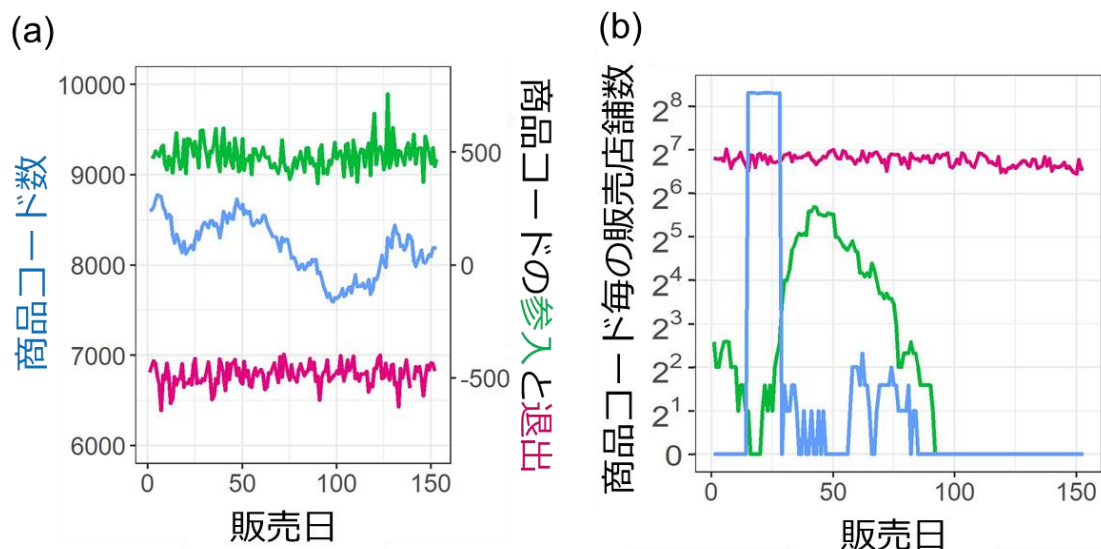


図 5.1 (a) 全 326 店舗での全商品コード数の推移。青線は日々の商品コードの全数、緑線は日々の商品コードの参入数、赤線は日々の商品コードの退出数。(b) 商品コード毎の販売店舗数の推移の例。赤線は販売店舗数が定常的な商品コード、青線は販売店舗数が急激に増減する商品コード、緑線は販売店舗数が山なりに推移する商品コード。

5.3 定常的な対数一様分布の存在

前節で述べた激しい商品の入れ替わりの中で、商品コード毎の販売店舗数の分布は、定常的に対数一様分布に従っていることを見出した。前述の図 5.1(b)は 3 つの商品コードの販売店舗数の推移例であるが、このような推移を全商品コードで調べ、特定の日々の各商品コードの販売店舗数を集計する。1 日目 (6/1)、62 日目 (8/1)、123 日目 (10/1) の各商品コードの販売店舗数の集計結果をヒストグラム表示したものが図 5.2(a)である。図 5.2(a)の横軸の店舗数は対数表示されており、ヒストグラムのビン、店舗数の最小値から最大値を \log 値で 9 等分した幅 (\log_2 で約 0.93) としている。縦軸の商品コード数は線形表示している。一般に、正の実現値 X の対数 $\log(X)$ が一様分布に従っている時、 X は対数一様分布に従っているとされる。1 日目、62 日目、123 日目のいずれの日のヒストグラムも、概ね一様な度数となっており、商品コード毎の店舗数は、時間によらず概ね対数一様分布に従っている。図 5.2(b)は図 5.2(a)のデータを累積度数分布で表示したものである。この累積度数分布の縦軸は CDF のように 1 に規格化せず個数で表示しており、分布形状に加えて、分

布毎の全数の違いを視覚化できるようにしている。図 5.2(b)の累積度数分布は横軸のみが対数表示されているため、グラフが直線状であることはデータが対数一様分布に従っていることを示している。赤線、緑線、青線は 1 日目、62 日目、123 日目の商品コード毎の店舗数の分布であり、対数一様分布は概ね定常的である。

図 5.3(a)は 1 日目の商品の販売店舗数の対数一様分布の推移である。赤線は 1 日目の商品の販売店舗数の分布である。緑線は 1 日目の商品のみの 62 日目の販売店舗数の分布、青は 1 日目の商品のみの 123 日目の販売店舗数の分布である。1 日目の対数一様分布は、商品の退出があっても概ね維持されていることがわかる。図 5.3(b)は 1 日目以降に参入した商品コードの販売店舗数の分布である。緑は 1 日目になかった商品のみの 62 日目の販売店舗数の分布、青は 1 日目になかった商品のみの 123 日目の販売店舗数の分布である。1 日目になかった参入商品のみでみた特定日の販売店舗数の分布も、対数一様分布となっている。

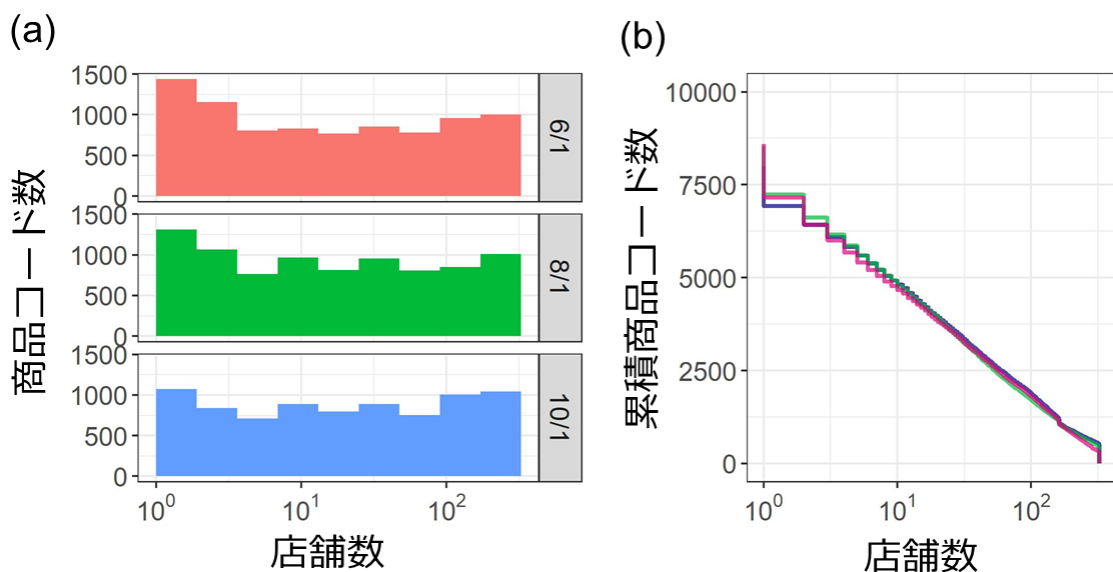


図 5.2 全商品コードの販売店舗数の分布。(a) ヒストグラム表示。(b) 累積度数表示。赤系列は 1 日目 (6/1)、緑系列は 62 日目 (8/1)、青系列は 123 日目 (10/1) の分布。

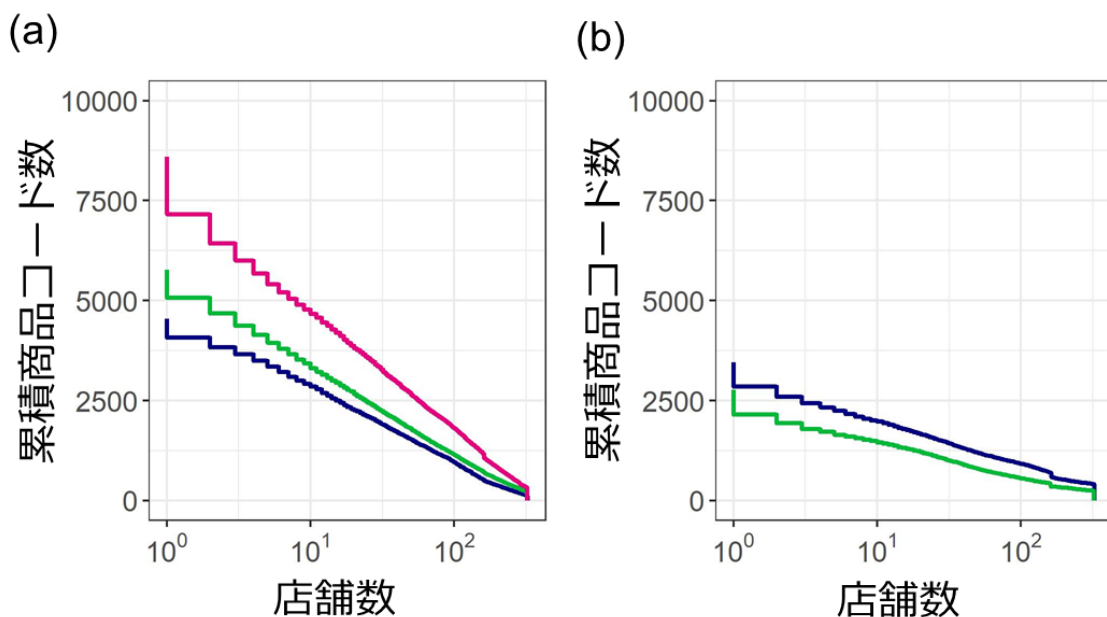


図 5.3 (a) 1 日目の対数一様分布の商品退出に伴う推移。赤は 1 日目の商品の販売店舗数の分布、緑は 1 日目の商品のみの 62 日目の販売店舗数の分布、青は 1 日目の商品のみの 123 日目の販売店舗数の分布。(b) 1 日目以降に参入した商品のみの販売店舗数の分布。緑は 1 日目になかった商品のみの 62 日目の販売店舗数の分布、青は 1 日目になかった商品のみの 123 日目の販売店舗数の分布。

5.4 対数一様分布の起源

対数一様分布の起源の考察のため、商品コード毎の販売店舗数の推移を詳細に分析した。図 5.1(b)に例示したように、商品コード毎の販売店舗数の推移は個別には複雑な推移をする。そこで、販売店舗数を 2 のべき乗 $[1, 2), [2, 2^2), [2^2, 2^3), \dots$ で Bin 分けし、商品コード毎の日々の販売店舗数がどの Bin からどの Bin に遷移したか、という形で推移の傾向性を確認する。全商品コードについて、全 153 日の Bin 間遷移を集計し、全体を 1 で規格化し \log_{10} をとった値をヒートマップ表示すると図 5.4(a)が得られる。横軸は遷移前の t 日目の販売店舗数の Bin、縦軸は遷移後の販売店舗数の Bin であり、 2^N は $[2^{N-1}, 2^N)$ の Bin を示している。横軸の 0 は参入前、縦軸の 0 は退出後を表す。このヒートマップより、いずれの Bin も隣接 Bin への遷移が多く、販売店舗数の推移は拡散的な振る舞いをしていることが分かる。

図 5.4(b)は、各 Bin から周辺 Bin への遷移量を色分けしてプロットしたものである。紫は 2^1 、青は 2^2 、薄青は 2^3 、薄緑は 2^4 、緑は 2^5 、黄は 2^6 、橙は 2^7 、茶は 2^8 、赤は 2^9 に対応する。

いずれの Bin についても、遷移前の Bin から離れた Bin への遷移量は小さい。遷移前の Bin に隣接する Bin とその隣の Bin までの遷移量を考慮し、最尤推定法により対数正規分布で各 Bin から周辺 Bin への遷移量を近似した。図 5.4(b)の各色の曲線は、そのようにして求めた対数正規分布を表している。2¹、2⁵、2⁸の Bin は代表として実線、その他の Bin は点線で示している。いずれの Bin の遷移量も対数正規分布で概ね近似できている。

図 5.4(c)は各 Bin の対数正規分布の平均と標準偏差である。赤点は標準偏差 $\log_2(\sigma)$ 、緑点は平均 $\log_2(\mu)$ である。概ね $\log_2(\mu) = 0$ 、 $\log_2(\sigma) = 0.5$ 程度であるが、Bin 毎に値が異なっている。赤点、緑点上の灰色線は2²から2⁸までの各データを用いた線形回帰結果である。ここで境界2¹、2⁹のデータはこの回帰分析では除外した。これらは分布の半分だけのデータから推定した平均と標準偏差であり、誤差が大きいと想定されるためである。得られた回帰式は、 t 日目の販売店舗数を $\log_2 x(t)$ とすると、 $\log_2(\sigma) = -0.089 \log_2 x(t) + 1.062$ 、 $\log_2(\mu) = 0.018 \log_2 x(t) - 0.182$ であった。図 5.4(c)の対数正規分布は、商品コード毎の日々の販売店舗数の推移の確率分布を表していると考えられる。対数正規分布を $\log_2 b(t)$ とすると $\log_2 x(t)$ は次の式に従うと仮定できる。

$$\log_2 x(t+1) = \log_2 x(t) + \log_2 b(t) \quad (5.1)$$

図 5.4(d)は各商品コードの $\log_2 b(t)$ の自己相関の箱ひげ図である。箱は四分位範囲 IQR、箱内の横線は中央値である。ひげの上端、下端は、箱の上端、下端から IQR の 1.5 倍の範囲の内での最大値と最小値を表している。縦軸 ± 0.196 の青の点線は、自己相関が 0 であるという帰無仮説の 95%信頼区間である。Lag 1 の 42%が青の点線を超えているものの、各 Lag の自己相関の絶対値は概ね青の点線より小さく、 $\log_2 b(t)$ は概ねランダムであると見做せる。これより、日々の商品コード毎の販売店舗数の推移は、式5.1のランダム乗算過程に従っていると仮定できる。 \log_2 を式5.1の両辺から除くと

$$x(t+1) = b(t)x(t) \quad (5.2)$$

となり、ある商品コードの $t+1$ 日目の販売店舗数 $x(t+1)$ は、 t 日目の販売店舗数 $x(t)$ に、確率的にランダムに変化する $b(t)$ が掛け合わされて得られる形となっている。

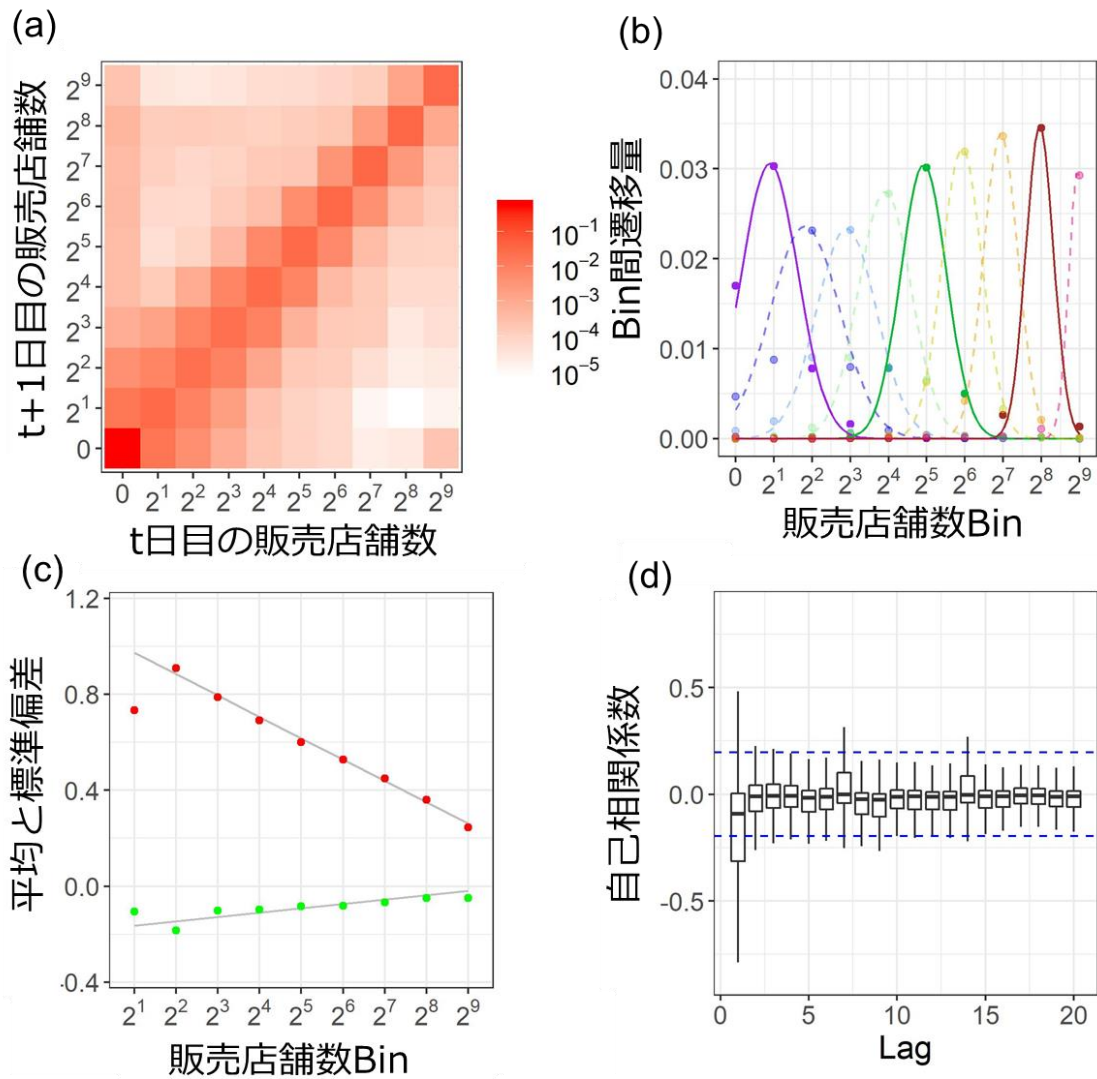


図 5.4 (a) 商品コード毎の日々の販売店舗数を2のべき乗 Bin で分けて集計した日々の Bin 間の遷移量。(b) 各 Bin から周辺 Bin への遷移量の対数正規分布近似。各点の色は遷移前の販売店舗数の2のべき乗 Bin に対応し、紫は 2^1 、青は 2^2 、薄青は 2^3 、薄緑は 2^4 、緑は 2^5 、黄は 2^6 、橙は 2^7 、茶は 2^8 、赤は 2^9 。各色の実線と点線は最尤推定法で求めた対数正規分布。(c) 各 Bin の対数正規分布のパラメータ推定結果。赤点は標準偏差 $\log_2(\sigma)$ 、緑点は平均 $\log_2(\mu)$ 。灰色線は線形回帰結果。(d) 乗算係数 $\log_2 b(t)$ の自己相関の箱ひげ図。青点線内は自己相関が0である帰無仮説の95%信頼区間。

シミュレーション検証（参入、退出なし）

ランダム乗算過程により、対数一様分布が現れることをシミュレーションにて確認する。シミュレーションのフローチャートを図 5.5 に示す。このシミュレーションでは、 $x(t)$ の時間発展は式 5.1 で決まるとしている。簡単のため、乗算係数 $\log_2 b(t)$ は Bin 依存なく、 $\log_2(\sigma) = 0.5$ 、 $\log_2(\mu) = 0$ の対数正規分布であるとした。また、商品の参入と退出は考慮しない。図 5.6(a)は、初期値を図 5.2(a)の 1 日目の対数一様分布から 500 個を非復元抽出したものとし、ランダム乗算を 100 ステップ実施した結果である。灰色線は初期分布、緑線は販売店舗数に上限下限を仮定しない時のランダム乗算結果である。緑線の分布は、横軸およそ 10^2 までは直線状となっているものの、より大きな値では値が発散する傾向が認められる。赤線は販売店舗数に上下限 $[1, 326]$ を仮定した時のランダム乗算結果である。最大値 326 は本研究の POS データの店舗数、最小値 1 は商品が存在できる最小の店舗数である。ランダム乗算の過程で、この上下限を超える値となった時には上下限の値で留まるとしている。このような上下限値を設定することで、対数一様分布が現れることが確認できる。

図 5.6 (b) は、初期値を対数一様分布からのサンプリングではなく、一定の値 100 が 500 個とし、販売店舗数の上限下限を $[1, 10000]$ とした時のランダム乗算結果である。乗算係数 $\log_2 b(t)$ は図 5.6(a)と同様に $\log_2(\sigma) = 0.5$ 、 $\log_2(\mu) = 0$ の対数正規分布としている。灰色線は初期分布、緑線は 100 ステップ後のランダム乗算結果、赤線は 1000 ステップ後のランダム乗算結果である。初期分布が対数一様分布でなく、また販売店舗数の上限下限が $[1, 10000]$ と広い場合にも、ランダム乗算のステップ数が 1000 などと十分であれば対数一様分布が現れる。

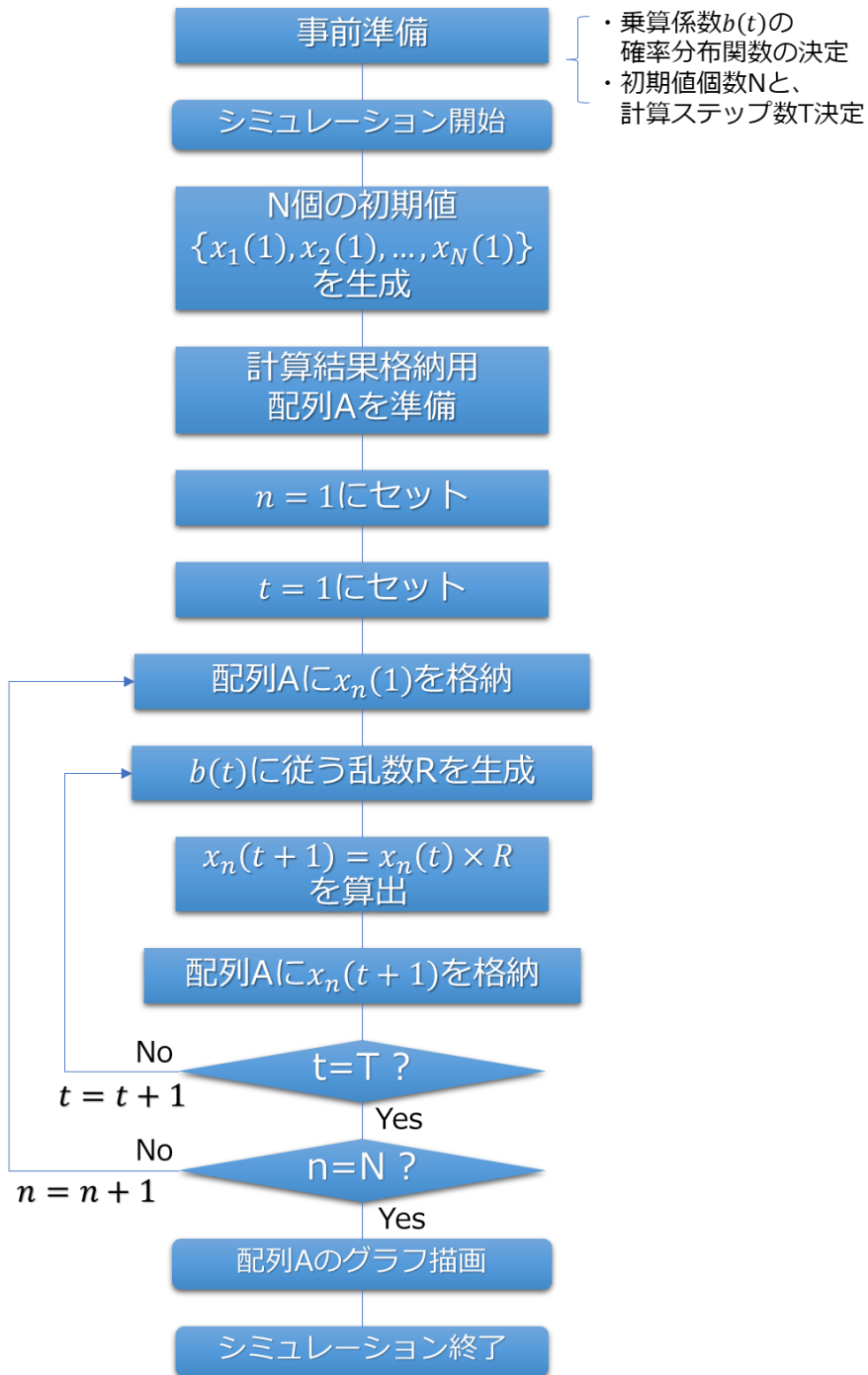


図 5.5 ランダム乗算シミュレーションのフローチャート

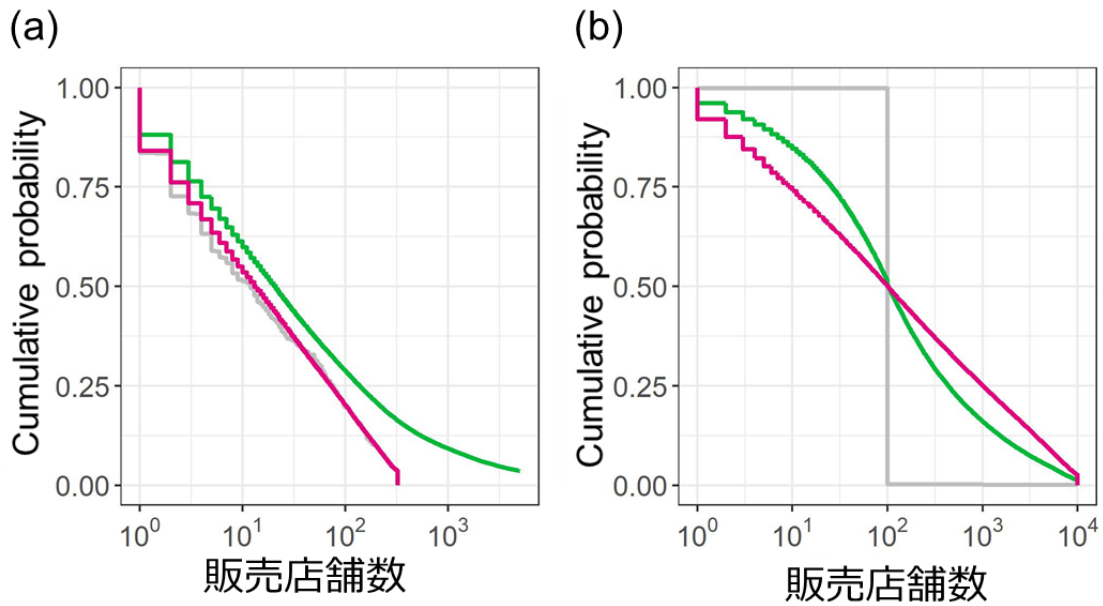


図 5.6 参入退出なしのランダム乗算過程のシミュレーション結果。(a) 乗算係数 $\log_2 b(t)$ は $\log_2(\sigma) = 0.5$ 、 $\log_2(\mu) = 0$ の対数正規分布とし、初期値として図 5.2(a) の 1 日目の対数一様分布から 500 個を非復元抽出して、ランダム乗算を 100 ステップ実施した結果。灰色線は初期分布、緑線は販売店舗数に上限下限を仮定しない時のランダム乗算結果。赤線は販売店舗数に上限下限 [1, 326] を仮定した時のランダム乗算結果。上限下限の境界外に値が出ることはできず境界の値で留まる。(b) 乗算係数 $\log_2 b(t)$ は $\log_2(\sigma) = 0.5$ 、 $\log_2(\mu) = 0$ の対数正規分布、初期値は 100 が 500 個、販売店舗数の上限下限を [1, 10000] とした時のランダム乗算結果。灰色は初期分布、緑は乗算ステップ 100、赤は乗算ステップ 1000 の時のランダム乗算結果。

商品の退出時間の見積もり

上記シミュレーションでは、商品の場合にみられる参入と退出を考慮していない。参入と退出を考慮したシミュレーションを行うため、ここでは商品が退出するまでの時間を見積もる。図 5.7 (a)は特定日にあった商品コードの数の推移である。青線は商品コード全数の推移、赤線、緑線、ネイビー線はそれぞれ 1 日目、62 日目、123 日目にあった商品コードの数の推移を表す。1 日目、62 日目、123 日目の商品コード数は、いずれも指数関数的に減少し、初期で早く減少し、以降遅く減少する傾向がみられる。そこで、初日 a_1 個の商品コードが t 日目に残る数 $N(t)$ を、次のような混合指数分布で近似する。

$$N(t) = a_1 \left\{ R_1 \exp\left(-\frac{t-1}{\mu_1}\right) + (1-R_1) \exp\left(-\frac{t-1}{\mu_2}\right) \right\} \quad (5.3)$$

ここで、 a_1 は 1 日目の商品コード数、 μ_1 と μ_2 は 2 つの指数分布の平均値、 R_1 は 2 つの指数分布の混合比率である。

これらの混合指数分布のパラメータは式5.3で得た商品コード数の減少カーブと、実データの商品コード数の減少カーブの RMSE が最小となるように見積もった。図 5.7 (a)の赤線、緑線、ネイビー線上の灰色線は、そのようにして得た式5.3による減少カーブである。混合指数分布のパラメータ (μ_1, μ_2, R_1) は 1 日目は $(\mu_1 = 8, \mu_2 = 255, R_1 = 0.17)$ 、62 日目は $(\mu_1 = 3, \mu_2 = 200, R_1 = 0.13)$ 、123 日目は $(\mu_1 = 5.5, \mu_2 = 290, R_1 = 0.14)$ であった。同様にして、実データの商品コード数の減少カーブを 25 個 (1 日目から 123 日目で 5 日おきに 1 日、6 日、11 日、...) を準備し、25 個の混合指数分布のパラメータを得た。その中央値は $(\mu_1 = 3, \mu_2 = 205, R_1 = 0.12)$ であり、四分位範囲は $(\mu_1 = 2, \mu_2 = 36.7, R_1 = 0.037)$ であった。商品コードの 12%が平均 3 日で退出し、残りは平均 205 日で退出すると解釈出来る。

商品の参入の見積もり

図 5.7 (b)は、図 5.4(a)から得た各 Bin での商品コードの参入量と退出量である。 2^4 から 2^8 では、概ね対数一様分布で参入していることが分かる。また、 2^4 から 2^8 と Bin 番号が大きくなるほど退出量が減少しており、実質的な参入量が増加している。一方、 2^1 から 2^3 、 2^9 においては参入量と退出量が同程度となっており、これらの Bin では実質的な参入量は少ないことが分かる。

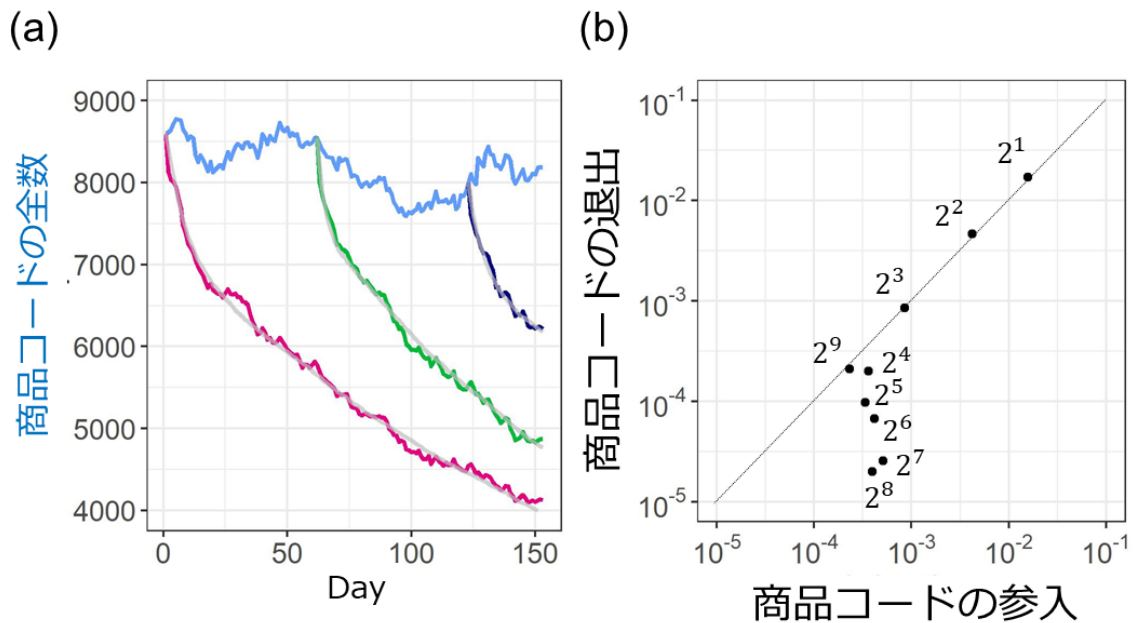


図 5.7 (a) 特定日にあった商品コードの数の推移。青線は商品コード全数の推移、赤線、緑線、ネイビー線はそれぞれ 1 日目、62 日目、123 日目にあった商品コードの数の推移。 (b) 各 Bin での商品コードの参入量と退出量。

参入と退出を考慮したシミュレーション

図 5.8 は、商品の参入と退出を考慮したランダム乗算シミュレーション結果である。灰色線は初期分布であり、図 5.2(a)の 1 日目の対数一様分布から 500 個を非復元抽出し、図 5.7 (b)の Bin 毎の参入量と退出量の比で間引いた参入分布とした。各初期値の寿命は、式5.3の混合指数分布 ($\mu_1 = 3, \mu_2 = 205, R_1 = 0.12$) からの乱数で決定した。具体的には、確率 0.12 で $\mu_1 = 3$ の指数分布、確率 0.88 で $\mu_2 = 205$ の指数分布から乱数を計 500 個取得し、それらの乱数の値を 500 個の初期値それぞれの寿命とした。ランダム乗算の途中で寿命を迎えた際には、以降のステップではその初期値による分布は消失し、結果の分布に寄与しないものとした。乗算係数 $\log_2 b(t)$ は、図 5.4(c)の回帰式より得た Bin 毎の平均と標準偏差を使用した。ランダム乗算の上下限は $[1, 326]$ に設定した。図 5.8 の青線は乗算ステップ 25、緑線は乗算ステップ 50、赤線は乗算ステップ 150 のランダム乗算結果である。赤線とほぼ重なっているネイビー線は、実データである図 5.2(a)の 1 日目の対数一様分布である。参入と退出を考慮したランダム乗算シミュレーションでも、対数一様分布が再現されることが確認できた。

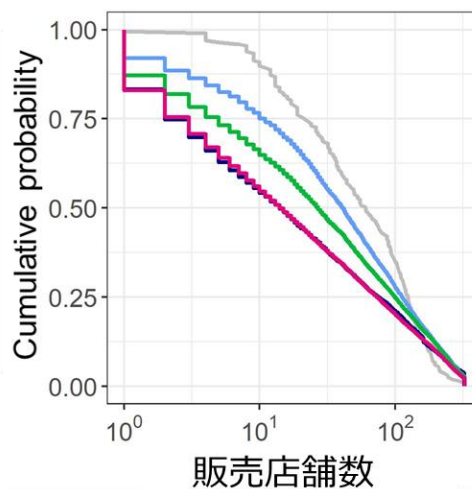


図 5.8 商品の参入と退出を考慮したランダム乗算シミュレーション結果。灰色線は初期分布で、図 5.2(a)の 1 日目の対数一様分布から 500 個を非復元抽出し、図 5.7 (b)の Bin 毎の参入量と退出量の比で間引いた参入分布。各初期値の寿命は、式5.3の混合指数分布 ($\mu_1 = 3, \mu_2 = 205, R_1 = 0.12$) からの乱数で決定。乗算係数 $\log_2 b(t)$ は図 5.4(c)の回帰式より得た Bin 毎の平均と標準偏差を使用。販売店舗数の上下限は $[1, 326]$ に設定。

商品の販売店舗数がランダム乗算過程に従う起源の考察

本章では、商品の販売店舗数がランダム乗算過程に従うこと、またそれにより対数一様分布が生成されることを数理的に明らかにした。一方、なぜ商品の販売店舗数がランダム乗算過程に従うのかについては明確になっていない。そのメカニズムを考察する。

一つ目の仮説として消費者側の効果を考える。ある商品の販売店舗数は、その商品に興味を持つ消費者の数に応じて変化すると仮定する。また、その消費者数の増減は、図 5.9 のような噂の伝播と、興味の減衰で決まると仮定する。噂の伝播の関連研究として、次の式 5.4 はミニブログ twitter におけるリツイート数、つまりある人が書いた記事が、その人の記事の読者であるフォロワに送付された後に、フォロワからまたそのフォロワへ記事が転送された回数を記述するものとして提案されたものである [157]。

$$n_m = \prod_{g=1}^m J_g N_0 \quad (5.4)$$

n_m はリツイート数、 J_g は対数正規分布の確率変数、 N_0 は最初に記事を書いた人のフォロワ数である。この式は、ある時点の記事の読者数は、その前の時点の記事の読者数に確率変数を掛けたもの、つまり乗算過程になることを表している。インターネット上の記事の伝播に限らず、口コミ等による商品の噂の伝播も、商品に興味をもつ人がある確率で別の人に興味を伝搬すると考えることで、同様な乗算過程になると考えられる。また、ある商品に興味をもつ消費者の減少についても、ある商品に興味をもつ消費者がある確率で興味を失うと考えることで乗算過程となる。従って、ある商品の販売店舗数がその商品に興味を持つ消費者の数に比例して増減するなら、販売店舗数は乗算過程で変化すると考えられる。

二つ目の仮説として販売者側の効果を考える。図 5.10 のように、ある商品の販売店舗数は、各地の店舗の経営指導をする本部の店舗指導員の影響を受けて変化すると仮定する。店舗指導員は、客の嗜好の変化や天候の情報などから商品の売れ筋の仮説を立て、各店舗の仕入れの精度を向上させる役割を担う。ここで店舗指導員は、商品が売れると想定される場合にはその商品の販売店舗数をある倍率 P 倍に増やし、売れていなければ $1/P$ 倍に減らすといった運用をすると仮定すると、商品の販売店舗数は乗算過程で推移することとなる。

このように商品の販売店舗数がランダム乗算過程に従うメカニズムはいくつか考えられるものの、考察の範囲に留まっており今後の研究が待たれる。

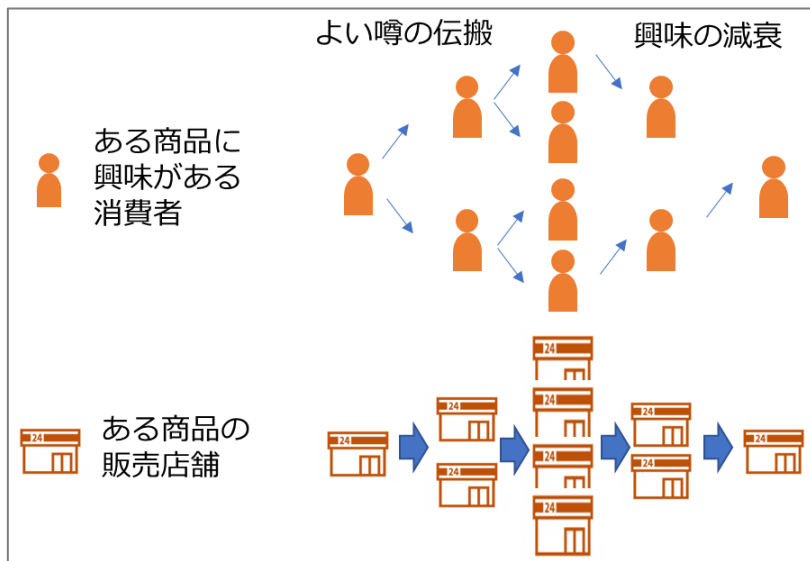


図 5.9 ある商品に興味を持つ消費者の乗算的な増減と、それに伴う販売店舗数の乗算的増減

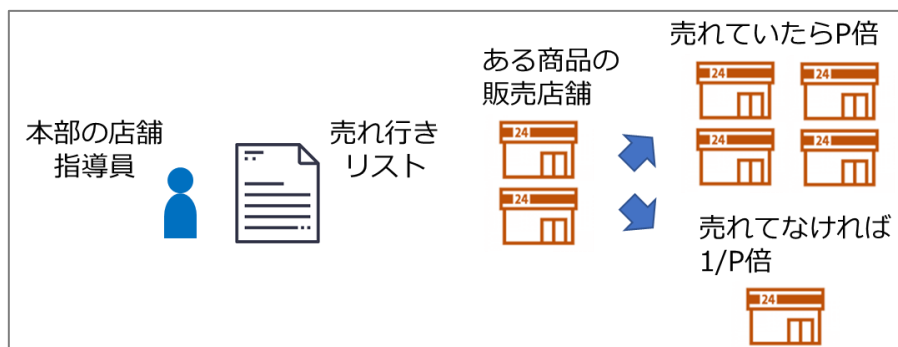


図 5.10 店舗指導員が売れ行きに応じて販売店舗数を乗算的に調整する様子

5.5 まとめ

本章では、商品コード毎の販売店舗数の変動と分布に関する数理モデルの開発を行った。POSデータの分析から、全326店舗の商品コード全数は日々概ね一定で8000個程度であるが、日々500個程度の商品コードの参入と退出があることを示し、また商品コード毎の販売店舗数の分布は定常的に対数一様分布に従うという経験則を見出した。商品コード毎の販売店舗数の時系列の解析から、商品コード毎の販売店舗数の変動がランダム乗算過程で近似できることを示した。数値シミュレーションによって、一般にランダム乗算過程に従って有限な区間内を変動することで対数一様分布が実現することを明らかにした。商品の退出時間は、長短の混合指数分布で表されることを示した。商品の参入と退出を考慮したランダム乗算過程に基づく数理モデルで対数一様分布を再現できることを確認した。

第6章 結言

本章では、本研究を総括するとともに今後の展望を述べる。

6.1 まとめ

本論文では、販売データの数理科学の研究を実施した。具体的には、コンビニエンスストアの詳細な販売データを解析し、販売時系列から需要を推定する手法、および、商品毎の販売店舗数の変動と分布に関する数理モデルの開発を行った。需要の推定においては、販売データの統計的性質であるテイラースケーリング則を考慮し、併せて非連続的に大きく変化する需要の推定を可能とする手法を粒子フィルタに導入することで、販売データと整合する適切な需要推定を可能とした。さらに、この需要推定法を応用し、消費期限のある商品に関して廃棄を低減しつつ利益の減少を最低限に留める方法を提案した。また、商品毎の販売店舗数が対数一様分布に従うという新たな統計的性質をデータから見出し、この分布の起源がランダム乗算過程に基づく数理モデルによって説明できることを明らかにした。

第1章では先行研究について述べ、本研究の目的を明らかにした。第2章では、本研究で用いた販売データの詳細を説明した。

第3章では、販売数時系列から需要を推定する手法として、テイラースケーリング則を考慮した非定常ポアソン時系列のパラメータ推定手法を確立した。具体的には、非定常時系列分析手法である粒子フィルタにおいて、観測モデルにテイラースケーリング則を考慮した尤度関数を導入し、さらに非連続的な需要の変動を推定するための非連続性の検知と推定値の補正の方法を開発し導入した。乱数で人工的に生成した時系列とPOSの販売時系列で提案手法を検証し、データと整合する適切な需要推定が可能となることを確認した。

第4章では、第3章で提案した時系列分析手法を拡張し、消費期限のある商品に対し、廃棄を低減しつつ利益の減少を最低限に留める方法を提案した。仕入数より需要が大きい時には売り切れによって販売データは需要分布の下限值となることを考慮し、観測上限で打ち切られたデータに対する既知の最尤推定法を粒子フィルタに導入した。さらに機会損失と廃棄損失の最適化問題の解として知られる新聞売り子問題の公式を拡張し、利益ロス

を見積もりながら所望の廃棄率を実現できる仕入数を決定する新しい手法を提案した。75%以上の販売日で廃棄がなされた食品の販売時系列 215 事例を用いた検証から、平均的に廃棄数を 23%に減少させながら利益を 140%に向上させる効果を確認した。さらに、複数店舗、複数商品の POS データを集計して販売数平均が大きい場合の効果検証を行い、テイラースケーリングの比例定数が小さい時、1%程度の利益ロスで廃棄を半減できるという知見を得た。

第5章では、商品コード毎の販売店舗数の変動と分布に関する数理モデルの開発を行った。販売データの分析から、商品コード毎の販売店舗数は定常的に対数一様分布に従うという経験則を見出した。商品コード毎の販売店舗数の時系列の解析から、商品コード毎の販売店舗数の変動がランダム乗算過程で近似できることを示し、数値シミュレーションによって、一般にランダム乗算過程に従って有限な区間内を変動することで対数一様分布が実現することを明らかにした。これらの結果を踏まえ、商品の入れ替わりを考慮したランダム乗算過程に基づく数理モデルによって販売店舗数の変動と分布の特性を再現することができることを確認した。

6.2 今後の展望

提案した非定常販売時系列のパラメータ推定手法の適用範囲に関して、第3章で述べたようにテイラースケーリング則に従うポアソン過程で記述される現象は、自然、社会において数多く確認されている。従って、提案手法は販売時系列に限らず、幅広い現象の非定常ポアソン時系列のパラメータ推定に活用できると見込まれる。また、この手法を拡張した廃棄低減の手法についても、コンビニエンスストアに限らず、食品製造業や、レストランといったサービス業など、幅広い業態の販売データにおいても適用できる可能性がある。一方、値引き販売の有無や、原価率の違いなど、業態ごとに前提となる条件が異なるため、提案手法の他業種展開に際しては、そういった前提条件の違いを考慮した効果検証が必要である。また提案したパラメータ推定手法においては、イベントは無相関で発生すると仮定しており、相関がある場合への対応に検討の余地がある。具体的には式3.2のテイラースケーリングの指数は無相関の場合の 1 を仮定しており、標準偏差は平均に比例するとみなしている。まとめ買いされる切手や発売日に多く売れる雑誌など、販売に相関がある場合には、テイラースケーリングの指数が 0.5 から 1 の間になることが知られている [28]。食品の場合にも、

流行やテレビ番組の影響などでそのような相関が生じる可能性がある。こういった相関がある場合への対応によって提案手法をより一般化させることが出来る。また、提案手法には販売の周期性、例えば週や季節の効果は組み込まれていない。それらの周期性の考慮によって需要推定精度を向上させられる可能性がある。予約や天気などの事前情報を考慮する拡張も有用であると考えられる。また、提案手法では、売り切れた場合に代替商品を選択する効果は考慮されていない。そういった商品間の相互作用の考慮もまた、利益を確保しながら廃棄を低減させる効果の向上に寄与すると考えられる。提案手法では需要は所与のものと想定しているが、価格の変更などによって需要を制御することも手法拡張の方向性として考えられる。

第 5 章で述べた対数一様分布の起源の数理モデルは、ランダム乗算過程という一般的な確率モデルを基礎としている。そのため、このモデルは、自然、社会の様々な現象での対数一様分布の考察に幅広く適用できる可能性がある。対数一様分布は、宇宙物理 [154]、地球科学 [155]、人間活動 [156]などで確認されており、これらの対数一様分布に我々のモデルが適用できるのかは興味深い研究課題である。一方、商品の販売店舗数がなぜランダム乗算過程に従うのかについては、第 5 章で消費者側の効果、販売者側の効果を論じたが、仮説の域を出ず今後の解明が待たれる。

謝辞

研究全般をご指導頂きました高安美佐子先生に感謝申し上げます。また高安秀樹先生に感謝申し上げます。お二人の先生方には、社会人として博士課程研究を始めるにあたり、研究室に入る前から親身にご相談に乗っていただきました。研究室に入った後にも、仕事との両立で平日に中々研究室に伺えない中、休日にも関わらずメールでご指導を頂いたり、時には研究室でお会い頂いたりしました。休日に投稿論文の修正やご指導を頂いたことも多く、そのような親身なご指導のお陰様で、仕事をしながらの博士課程研究を進めることが出来ました。研究テーマにつきましても、私にとって本当興味深いもので、充実した研究をさせて頂きましたことに感謝申し上げます。

本論文の審査をして頂きました出口弘先生、三宅美博先生、青西亨先生、小野功先生にはお忙しい中、本論文の原稿をご確認頂き、また審査では的確なご指摘を頂きましてありがとうございました。お陰様で本論文を改善することが出来ました。

高安研究室の助教の金澤輝代士さん、田村光太郎さん、尾崎順一さん、由良嘉啓さん、講師の山田健太さんにも研究内容や、研究生活の面で多くのご助言をいただきました。研究室のメンバーの皆様、特に小林祐一朗さん、末重拓己さん、Arthur Matsuo さん、志田洋平さんには、東工大での生活や授業、研究に関する様々なことで相談に乗って頂き感謝しております。

高安研究室秘書の鳥井麻希さん、赤羽晶子さんにも、研究生活でのご相談に乗って頂いたり、多くのご配慮を頂きました。平日研究室に伺えない中、私の行うべき書類の作成や提出を行って頂いたことも度々ありました。どうもありがとうございました。

多くの方々のご尽力やご配慮のお陰様で本研究を進めることが出来ました。改めて感謝申し上げます。ありがとうございました。

参考文献

- [1] 鷲尾 和紀, ONE to ONE マーケティングとリレーションシップ・マーケティング, 高千穂論叢, 49(1), 2014-5, 273-311..
- [2] 小野潔, 金融業におけるデータマイニングの応用 保険解約の防止分析, オペレーションズ・リサーチ, 45(5), 2000-05-01, 233-238..
- [3] 南 皓輔, 今村 岳, 柴 弘太, 吉川 元起, 嗅覚 IoT センサ「膜型表面応力センサ (MSS)」とヘルスケア・医療分野への応用可能性, 電気化学, 86, 2018, 123-128..
- [4] 西川 凜, IoT デバイスを用いた酪牛の行動検知・認識, 組込みシステムシンポジウム 2018 論文集, 2018, 40-43..
- [5] アズビル株式会社, オンライン異常予兆検知システム BiG EYES, 計測と制御, 55(7), 2016, 627-628..
- [6] 佐野 慎一, 挟間 洋平, 澤田 宗基, IoT による発電機保全データ収集と故障予兆診断の取り組み, 明電時報, 4, 2016, 52-56..
- [7] Mantegna, R.N.; Stanley, H.E.; Nature, 1995, 376, 46-48..
- [8] Yura, Y.; Ohnishi, T.; Yamada, K.; Takayasu, T.; Takayasu, M. Replication of Non-Trivial Directional Motion in Multi-Scales Observed by the Runs Test, International Journal of Modern Physics: Conference Series 2012, 16, pp136-148..
- [9] Takayasu, M; Mizuno, T; Takayasu, H. Potentials force observed in market dynamics, Physica A 2006, 370, 91-97..
- [10] Watanabe, K.; Takayasu, H.; Takayasu, M. Random walker in temporally deforming higher-order potential forces observed in a financial crisis, Physical Review E, 2009, 80, 056110.
- [11] 由良 嘉啓, 高安 美佐子, 外国為替市場のシミュレーション, シミュレーション 33(4), 2014-12, 254-257..
- [12] マーク・ブキャナン, 市場は物理法則で動く, 白揚社, 2015; pp. 386..
- [13] Miura, W.; Takayasu, H.; Takayasu, M. Effect of Coagulation of Nodes in an Evolving Complex Network. Phys. Rev. Lett. 2012, 108, 168701..
- [14] Tamura, K.; Miura, W.; Takayasu, M.; Takayasu, H.; Kitajima, S.; Goto, H. Estimation of Flux Between Interacting Nodes on Huge Inter-Firm Networks, International Journal of Modern Physics: Conference Series, 2012, 16, 93-104..
- [15] Tamura, K.; Takayasu, H.; Takayasu, M. Diffusion-localization transition caused by

- nonlinear transport on complex networks. *Sci Rep.* 2018, 8(1), 1-7..
- [16] 高安 美佐子, 三浦 航, 田村 光太郎, 企業取引ネットワークの形成と特性, システム／制御／情報 56(10), 2012, 517-522..
- [17] 後藤 隼人, 高安 美佐子, 企業間取引ネットワークとそのシミュレーション : ネットワークの成長過程, 輸送問題, および頑強性評価, オペレーションズ・リサーチ 60(8), 2015, 456-462..
- [18] Groot, R.D. Lévy distribution and long correlation times in supermarket sales, *Physica A*, 2005, 353, 501–514..
- [19] Groot, R.D. Consumers don't play dice, influence of social networks and advertisements, *Physica A*, 2006, 363, 446–458..
- [20] Sornette, D; Deschâtres, F; Gilbert, T; Ageon Y. Endogenous versus exogenous shocks in complex networks: An empirical test using book sale rankings. *Phys Rev Lett.* 2004, 93(22), 1-4..
- [21] Mizuno, T. Power law of customers' expenditures in convenience stores. *J Phys Soc Japan*, 2008, 77(3), 2-3..
- [22] Ueno, H; Watanabe, T; Takayasu, H; Takayasu, M. Collective purchase behavior toward retail price changes. *Phys A Stat Mech its Appl.* 2011, 390(3), 499-504..
- [23] Hisano, R.; Mizuno, T. Sales distribution of consumer electronics, *Phys A Stat Mech its Appl*, 2010, 390(2), 309-318..
- [24] Mizuno, T.; Watanabe, T. A statistical analysis of product prices in online markets, *Eur Phys J B*, 2010, 76(4), 501-505..
- [25] Mizuno, T; Toriyama, M; Terano, T; Takayasu, M. Pareto law of the expenditure of a person in convenience stores. *Phys A Stat Mech its Appl.* 2008, 387(15), 3931-3935..
- [26] Montroll, EW. On the entropy function in sociotechnical systems. *Proc Natl Acad Sci U S A.* 1981, 78, 12II, 7839-7843..
- [27] Mizuno, T; Takayasu, M. The statistical relationship between product life cycle and repeat purchase behavior in convenience stores. *Prog Theor Phys Suppl.* 2009, 179, 71-79..
- [28] Fukunaga, G.; Takayasu, H.; Takayasu, M. Property of Fluctuations of Sales Quantities by Product Category.
- [29] Doganis, P.; Alexandridis, A.; Patrinos, P.; Sarimveis, H. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *J. Food Eng.* 2006, 75, 196–204..

- [30] Huber, J.; Gossmann, A.; Stuckenschmidt, H. Cluster-based hierarchical demand forecasting for perishable goods. *Expert Syst. Appl.* 2017, 76, 140–151..
- [31] Yang, C.L.; Sutrisno, H. Short-Term Sales Forecast of Perishable Goods for Franchise Business. In *Proceedings of the 2018 10th International Conference on Knowledge and Smart Technology: Cybernetics in the Next Decades*, Chiang Mai, Thailand, 31 January–3.
- [32] Lee, W.I.; Chen, C.W.; Chen, K.H.; Chen, T.H.; Liu, C.C. A comparative study on the forecast of fresh food sales using logistic regression, moving average and bpnn methods. *J. Mar. Sci. Technol.* 2012, 20, 142–152..
- [33] Chen, C.Y.; Lee, W.I.; Kuo, H.M.; Chen, C.W.; Chen, K.H. The study of a forecasting sales model for fresh food. *Expert Syst. Appl.* 2010, 37, 7696–7702..
- [34] Kiil, K.; Hvolby, H.H.; Fraser, K.; Dreyer, H.; Strandhagen, J.O. Automatic replenishment of perishables in grocery retailing. *Br. Food J.* 2018, 120, 2033–2046..
- [35] Hunter, J.S. The Exponentially Weighted Moving Average. *J. Qual. Technol.* 1986, 18, 203–210..
- [36] Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1976..
- [37] Reyman, G. State reduction in a dependent demand inventory model given by a time series. *Eur. J. Oper. Res.* 1989, 41, 174–180..
- [38] Graves, S.C. A single-item inventory model for a nonstationary demand process. *Manuf. Serv. Oper. Manag.* 1999, 1, 174–180..
- [39] Carrizosa, E.; Olivares-Nadal, A.V.; Ramírez-Cobo, P. Robust newsvendor problem with autoregressive demand. *Comput. Oper. Res.* 2016, 68, 123–133..
- [40] Zhang, G.P. *Neural Networks for Time-Series Forecasting*, *Handbook of Natural Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 461–477..
- [41] See, C.T.; Sim, M. Robust Approximation to Multiperiod Inventory Management. *Oper. Res.* 2010, 58, 583–594..
- [42] Ban, G.; Rudin, C. The Big Data Newsvendor: Practical Insights from Machine Learning. *Oper. Res.* 2019, 67, 90–108..
- [43] Winkelmann, R. *Econometric Analysis of Count Data*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 63–126..
- [44] Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman and Hall: London, UK, 1990..

- [45] Harvey, A. Forecasting with Unobserved Components Time Series Models, Handbook of Economic Forecasting; Elsevier: Amsterdam, The Netherlands, 2006..
- [46] 山口類, 土屋映子, 樋口知之, 状態空間モデルを用いた飲食店売上の要因分解, オペレーションズ・リサーチ, 49(5), 2004-05-01, 316-324.
- [47] 本橋 永至, 樋口 知之, 市場構造の変化を考慮したブランド選択モデルによる購買履歴データの解析, マーケティング・サイエンス, 2013, 21, 37-59..
- [48] 青柳 憲治, 佐藤 忠彦, 3 階層多変量状態空間モデリングによる動的市場反応形成メカニズムの解明, Transactions of the Operations Research Society of Japan. 2015, 58, 70-100..
- [49] F.M.Bass, N.Bruce, S.Majumdar, B.P.S.Murthi, Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising-Sales Relationship, Marketing Science, 2007, 26, 179-195..
- [50] 佐藤 忠彦, 樋口 知之, POS データの時系列モデリングによる知識発見, 人工知能学会論文誌, 2007, 22, 200-208..
- [51] M.Lachaab, A.Ansari, K.Jedidi, A.Trabelsi, Modeling preference evolution in discrete choice models: A Bayesian state-space approach, Quantitative Marketing and Economics, 2006, 4, 57-81..
- [52] Sato,T.; Higuchi,T.; KitagawaG. Statistical Inference using Stochastic Switching Models for the Discrimination of Unobserved Display Promotion form POS Data., Marketing Letters, 2004, 15, 37-60..
- [53] Q.Ma, W.Zhang, Public Mood and Consumption Choices: Evidence from Sales of Sony Cameras on Taobao PLoS ONE, 2015, 10(4), e0123129..
- [54] 石垣司, 竹中毅, 木村陽一, 日常購買行動に関する大規模データの融合による顧客行動予測システム, 人工知能学会論文誌 26(6), 2011, 670-681.
- [55] Porteus, E.L. Stochastic Inventory Theory. Handbooks in Operations Research and Management Science; Elsevier: Amsterdam, The Netherlands, 1990; Volume 2, pp. 605-652..
- [56] Khouja, M. The single-period (news-vendor) problem: Literature review and suggestions for future research. Omega 1999, 27, 537-553..
- [57] Qin, Y.; Wang, R.; Vakharia, A.J.; Chen, Y.; Seref, M.M.H. The newsvendor problem: Review and directions for future research. Eur. J. Oper. Res. 2011, 213, 361-374..
- [58] Congdon, P. Bayesian Models for Categorical Data, 2005, Wiley..
- [59] Scarf, H.E. Bayes Solutions of the Statistical Inventory Problem. Ann. Math. Statist.

- 1959, 30, 490–508..
- [60] Scarf, H.E. Some remarks on bayes solutions to the inventory problem. *Naval Res. Logist. Quart.* 1960,7, 591–596..
- [61] Azoury, K.S. Bayes Solution to Dynamic Inventory Models under Unknown Demand Distribution. *Manag. Sci.* 1985, 31, 1150–1160..
- [62] Conrad, S.A. Sales Data and the Estimation of Demand. *Oper. Res. Q.* 1976, 27, 123–127..
- [63] Wecker, W.E. Predicting Demand from Sales Data in the Presence of Stockouts. *Manag. Sci.* 1978, 24, 1043–1054..
- [64] Braden, D.J.; Freimer, M. Informational Dynamics of Censored Observations. *Manag. Sci.* 1991, 37, 1390–1404..
- [65] Lariviere, M.A.; Porteus, E.L. Stalking Information: Bayesian Inventory Management with Unobserved Lost Sales. *Manag. Sci.* 1999, 45, 346–363..
- [66] Lovejoy, W.S. Myopic Policies for Some Inventory Models with Uncertain Demand Distributions. *Manag. Sci.* 1990, 36, 724–738..
- [67] Chen, L. Bounds and Heuristics for Optimal Bayesian Inventory Control with Unobserved Lost Sales. *Oper. Res.* 2010, 58, 396–413..
- [68] Ding, X.; Puterman, M.L.; Bisi, A. The Censored Newsvendor and the Optimal Acquisition of Information. *Oper. Res.* 2002, 50, 517–527..
- [69] Lu, X.; Song, J.S.; Zhu, K. Analysis of Perishable-Inventory Systems with Censored Demand Data. *Oper. Res.* 2008, 56, 1034–1038..
- [70] Besbes, O.; Chaneton, J.; Moallemi, C.C. The Exploration-Exploitation Tradeoff in the Newsvendor Problem. *SSRN Electron. J.* 2017, 14–61..
- [71] Burnetas, A.N.; Smith, C.E. Adaptive Ordering and Pricing for Perishable Products. *Oper. Res.* 2000, 48, 436–443..
- [72] Godfrey, G.A.; Powell, W.B. An Adaptive, Distribution-Free Algorithm for the Newsvendor Problem with Censored Demands, with Applications to Inventory and Distribution. *Manag. Sci.* 2001, 47, 1101–1112..
- [73] Huh, W.T.; Rusmevichientong, P. A Nonparametric Asymptotic Analysis of Inventory Planning with Censored Demand. *Math. Oper. Res.* 2009, 34, 103–123..
- [74] Wang, Z.F.; Mersereau, A.J. Bayesian Inventory Management with Potential Change-Points in Demand. *Prod. Oper. Manag.* 2017, 26, 341–359..

- [75] Chen, L.; Mersereau, A.J. *Analytics for Operational Visibility in the Retail Store: The Cases of Censored Demand and Inventory Record Inaccuracy*. International Series in Operations Research & Management Science; Springer: Berlin/Heidelberg, Germany, 2015;.
- [76] *World Population Prospects the 2017 Revision*; United Nations: New York, NY, USA, 2017..
- [77] *Food Wastage Footprint—Impacts on Natural Resources*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2013..
- [78] *Global Food Security Index 2014 Special Report: Food Loss and Its Intersection with Food Security*; The Economist Intelligence Unit: London, UK, 2014..
- [79] Lindgren, E.; Harris, F.; Dangour, A.D.; Gasparatos, A.; Hiramatsu, M.; Javadi, F.; Loken, B.; Murakami, T.; Scheelbeek, P.; Haines, A. Sustainable food systems—A health perspective. *Sustain. Sci.* 2018, 13, 1505–1517..
- [80] *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015..
- [81] FUSIONS. Estimates of European Food Waste Levels. Available online: <http://www.fusions.org/phocadownload/Publications/Estimates%20of%20European%20food%20waste%20levels.pdf>.
- [82] Goldenberg, S. From Field to Fork: the Six Stages of Wasting Food. Available online: <https://www.theguardian.com/environment/2016/jul/14/from-field-to-fork-the-six-stages-of-wasting-food>.
- [83] Estimates of Food Surplus and Waste Arisings in the UK. Available online: http://www.wrap.org.uk/sites/files/wrap/Estimates_%20in_the_UK_Jan17.pdf.
- [84] *Global Food Losses and FoodWaste*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2011..
- [85] Katajajuuri, J.; Silvennoinen, K.; Hartikainen, H.; Heikkilä, L.; Reinikainen, A. Food waste in the Finnish food chain. *J. Clean. Prod.* 2014, 73, 322–329..
- [86] Brancoli, P.; Rousta, K.; Bolton, K. Life cycle assessment of supermarket food waste. *Resour. Conserv. Recycl.* 2017, 118, 39–46..
- [87] Falcone, P.M.; Imbert, E. *Bringing a Sharing Economy Approach into the Food Sector: The Potential of Food Sharing for Reducing Food Waste*; Springer: Berlin/Heidelberg, Germany, 2017..

- [88] Morone, P.; Falcone, P.M.; Imbert, E.; Morone, A. Does food sharing lead to food waste reduction? An experimental analysis to assess challenges and opportunities of a new consumption model. *J. Clean. Prod.* 2018, 185, 749–760..
- [89] Leftoverswap. Available online: <http://leftoverswap.com/>.
- [90] Foodsharing. Available online: <https://foodsharing.de/>.
- [91] Eatwith. Available online: <https://www.eatwith.com/>.
- [92] Feastly. Available online: <https://eatfeastly.com/>.
- [93] Galanakis, C. *Sustainable Food Systems from Agriculture to Industry*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 371–399..
- [94] Lin, C.S.K.; Koutinas, A.A.; Stamatelatou, K.; Mubofu, E.B.; Matharu, A.S.; Kopsahelis, N.; Pfaltzgraff, L.A.; Clark, J.H.; Papanikolaou, S.; Kwan, T.H.; et al. Current and future trends in food waste valorization for the production of chemicals, material.
- [95] Love Food HateWaste. Available online: <https://www.lovefoodhatewaste.com/node/2343>.
- [96] I Love Food. Available online: <https://ivaluefood.com/>.
- [97] FareShare. Fighting Hunger, Tackling Food Waste. Available online: <http://www.fareshare.org.uk>.
- [98] Filimonau, V.; Gherbin, A. An exploratory study of food waste management practices in the UK grocery retail sector. *J. Clean. Prod.* 2017, 167, 1184–1194..
- [99] Food Cloud. Available online: <http://food.cloud/>.
- [100] ASDA and the Environment. Available online: <https://www.asda.com/environment>.
- [101] Food Labelling Terms. Available online: <http://www.nhs.uk/Livewell/Goodfood/Pages/food-labellingterms.aspx>.
- [102] Heikkilä, L.; Reinikainen, A.; Katajajuuri, J.M.; Silvennoinen, K.; Hartikainen, H. Elements affecting food waste in the food service sector. *Waste Manag.* 2016, 56, 446–453..
- [103] Martin-Rios, C.; Demen-Meier, C.; Gössling, S.; Cornuz, C. Food waste management innovations in the foodservice industry. *Waste Manag.* 2018, 79, 196–206.
- [104] Tasaki, T.; Yamakawa, H. An estimation of the effectiveness of waste prevention by using point-of-sales (POS) data—The case of refills for shampoo and hair conditioner in Japan. *Resour. Conserv. Recycl.* 2011, 57, 61–66..
- [105] Kendal, W.S. A frequency distribution for the number of hematogenous organ

- metastases. *J. Theor. Biol.* 2002, 217, 203–218..
- [106] Eisler, Z.; Kertész, J. Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. *Phys. Rev. E.* 2006, 73, 046109..
- [107] De Menezes, M.A.; Barabási, A.-L. Fluctuations in network dynamics. *Phys. Rev. Lett.* 2004, 92, 028701..
- [108] Sano, Y.; Yamada, K.; Watanabe, H.; Takayasu, H.; Takayasu, M. Empirical analysis of collective human behavior for extraordinary events in the blogosphere. *Phys. Rev. E.* 2013, 87, 012805..
- [109] 阿部 誠, 近藤 文代, マーケティングの科学—POS データの解析, 朝倉書店..
- [110] 【IT 部門が教える】今さら聞けない POS システムの全知識と成功の条件, https://it-koala.com/pos_system-548#POS-7..
- [111] Seven & i Holdings Co., Ltd. Corporate Outline 2011. Available online: https://www.7andi.com/library/dbps_data/_template/_res/en/ir/library/co/pdf/2011_07.pdf.
- [112] Sakoda, G.; Takayasu, H.; Takayasu, M. Tracking Poisson Parameter for Non-Stationary Discontinuous Time Series with Taylor’s Abnormal Fluctuation Scaling. *Stats.* 2019, 2(1), 55-69..
- [113] Taylor, L.R. Aggregation, variance and the mean. *Nature* 1961, 189, 732–735..
- [114] Eisler, Z.; Bartos, I.; Kertész, J. Fluctuation scaling in complex systems: Taylor’s law and beyond. *Adv. Phys.* 2008, 57, 89..
- [115] Meloni, S.; Gomez-Gardeñes, J.; Latora, V.; Moreno, Y. Scaling Breakdown in Flow Fluctuations on Complex Networks. *Phys. Rev. Lett.* 2008, 100, 208701.
- [116] Draper, N.R.; Smith, H. *Applied Regression Analysis*; Wiley: Hoboken, NJ, USA, 1998; pp. 505–565.
- [117] Ritz, C.; Streibig, J.C. *Nonlinear Regression with R*, Springer, 2008..
- [118] Shi, P.; Ratkowsky, D.A.; Wang, N.; et al. Comparison of five methods for parameter estimation under Taylor’s power law. *Ecol Complex.* 2017, 32, 121-130..
- [119] 樋口 知之, 予測にいかず統計モデリングの基本—ベイズ統計入門から応用まで, 2011, 講談社..
- [120] Yura, Y.; Takayasu, H.; Nakamura, K.; Takayasu, M., *J. Stat. Comput. Simul.* 2014, 84, 2073–2090.
- [121] Cherry, S.R.; Sorenson, J.A.; Phelps, M.E. *Physics in Nuclear Medicine*, 4th ed.;

- Elsevier, 2012; pp. 126–128..
- [122] 岩波データサイエンス刊行委員会, 岩波データサイエンス Vol. 6, 岩波書店..
- [123] Kitagawa, G. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models., *J. Comput. Graph. Stat.* 1996, 5, 1–25.
- [124] Gordon, N.J.; Salmond, D.J.; Smith, A.F. Novel approach to nonlinear/non-Gaussian Bayesian state, estimation. *IEE Proc. F-Radar Signal Process.* 1993, 140, 107–113..
- [125] Walker, A.J. New fast method for generating discrete random numbers with arbitrary frequency distributions, *Electronics Letters*, 1974, 10 (8), 127..
- [126] Spector, P. *Data Manipulation with R*, Springer, 2008..
- [127] Luc, D. *Non-Uniform Random Variate Generation*, Springer, 1986, pp.27–36..
- [128] Box, G.E.P.; Muller, M.E. A note on the generation of random normal deviates, *Annals Math. Stat.*, 1958, 29, 610-611..
- [129] Sakoda, G.; Takayasu, H.; Takayasu, M. Data science solutions for retail strategy to reduce waste keeping high profit. *Sustain.* 2019, 11(13), 1-30..
- [130] Sakoda, G.; Takayasu, H.; Takayasu, M. Retail Strategy to Reduce Waste Keeping High Profit Utilizing Taylor's Law in Point-of-Sales Data, *International Journal of Physical and Mathematical Sciences*, 2019, 13(12), 213-218..
- [131] Lee, E.T.; Wang, J.W. *Statistical Methods for Survival Data Analysis*; WILEY: Hoboken, NJ, USA, 2013; pp. 133–205..
- [132] Huh, W.T.; Levi, R.; Rusmevichientong, P.; Orlin, J.B. Adaptive Data-Driven Inventory Control with Censored Demand Based on Kaplan-Meier Estimator. *Oper. Res.* 2011, 59, 929–941..
- [133] Mazzucato, M.; Semieniuk, G. Financing renewable energy: Who is financing what and why it matters. *Technol. Forecast. Soc. Chang.* 2018, 127, 8–22..
- [134] Falcone, P.M.; Sica, E. Assessing the opportunities and challenges of green finance in Italy: An analysis of the biomass production sector. *Sustainability* 2019, 11, 517..
- [135] Sakoda, G.; Takayasu, H.; Takayasu, M. Metabolic dynamics of ecosystems realizing steady log-uniform distributions: The case of commodities in shops. *Entropy.* 2020, 22(3), 267..
- [136] Smoluchowski, M.V. *Zeitschrift für Physik* 1916, 17, 585..
- [137] Takayasu, H. Steady-state distribution of generalized aggregation system with injection. *Phys. Rev. Lett.* 1989, 63, 2563..

- [138] Takayasu, M. Universal power law observed in an exponentially growing particle system. *Phys. Rev. A.* 1992, 46, 782..
- [139] Takayasu, H.; Taguchi, Y-h. Non-Gaussian distribution in random advection dynamics. *Phys. Rev. Lett.* 1993, 70, 782..
- [140] Takayasu, M.; Takayasu, H.; Taguchi, Y-h. Non-Gaussian distribution in random transport dynamics. *Int. J. Mod. Phys. B* 1994, 8, 3887..
- [141] Krapivsky, P.L; Redner, S. Transitional aggregation kinetics in dry and damp environments. *Phys. Rev. E.* 1996, 54, 3553..
- [142] Friedlander, S.K. Smoke, dust, and haze: fundamentals of aerosol dynamics 2nd ed;Oxford University Press: New York, 2000..
- [143] Bak, P.; Tang, C.;Wiesenfeld, K. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.* 1987, 59, 381..
- [144] Lu, E.T.; Hamilton, R.J. Avalanches and the Distribution of Solar Flares. *Astrophys. J.* 1991, 380, 89..
- [145] Olami, Z.; Feder, H.J.S.; Christensen, K. Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Phys. Rev. Lett.* 1992, 68, 1244..
- [146] Mizuno, T.; Kurihara, S.; Takayasu, M; Takayasu, H. Analysis of high-resolution foreign exchange data of USD-JPY for 13 years. *Phys. A.* 2003, 324, 296..
- [147] Okuyama, K.; Takayasu, M.; Takayasu, H. Zipf's law in income distribution of companies. *Phys. A.* 1999, 269, 125..
- [148] Yule, G.U. II.—A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. *Phil. Trans. R. Soc. London B.* 1925, 213, 21..
- [149] Simon, H.A. On a Class of Skew Distribution Functions. *Biometrika.* 1955, 42, 425..
- [150] Barabási, R.; Albert, A. Emergence of Scaling in Random Networks. *Science.* 1999, 286, 509..
- [151] Feller,W. An Introduction to Probability Theory and Its Applications 2nd ed vol 2; Wiley: New York, 1971..
- [152] Levy, M; Solomon, S. Power laws are logarithmic Boltzmann laws. *Int. J. Mod. Phys. C.* 1996, 7, 595..
- [153] Takayasu, H; Sato, A.H.; Takayasu,M. Stable Infinite Variance Fluctuations in Randomly Amplified Langevin Systems. *Phys. Rev. Lett.* 1997, 79, 966..

- [154] Connelley, M.S.; Reipurth, B.; Tokunaga, A.T. The Evolution of the Multiplicity of Embedded Protostars II: Binary Separation Distribution & Analysis. *Astron. J.* 2008, 135, 2526..
- [155] Rabeaud, O; Royer, J.J; Jébrak, M; Cheilletz, A. Log-uniform distribution of gold deposits along major Archean fault zones. *Miner. Depos.* 2013, 48, 817..
- [156] Tsafrir, D.; Etsion, Y; Feitelson, D.G. Modeling User Runtime Estimates. *Job Sched. Strateg. Parallel Process.* 2005, 3834, 1..
- [157] Kawamoto; T., A stochastic model of tweet diffusion on the Twitter network, *Phys A Stat Mech its Appl.* 2013, 392(16), 3470-3475..