

平成 12 年度  
学士学位論文

# POS における標本自己相関に関する研究

A Study on Sample Auto Correlation in POS Data

1010463 湯藤 佳忠

指導教員 Ruck Thawonmas

2001 年 2 月 5 日

高知工科大学 情報システム工学科

# 要 旨

## POS における標本自己相関に関する研究

湯藤 佳忠

流通業において、在庫需要に合わせた最適な発注判断は、企業の経営や競争力を左右する重要な業務である。従来は人間の経験則に依存していたため、在庫過大による経営の圧迫や、在庫過小による販売機会の逸失といった問題があった。この問題を解決するために、企業と本大学による共同プロジェクト「ニューラルネットによる商品発注支援システムに関する研究」を立案した。本論分では同プロジェクトの初期研究結果である、POS データの時系列解析及び標本自己相関による販売傾向の分析結果を報告する。

**キーワード** 時系列解析, 標本自己相関

# Abstract

## A Study on Sample Auto Correlation in POS Data

Yoshitada YUTO

For distribution industry, efficient judgment of order is an important business issue that determines the management and competitive power of a company. Until now, since ordering was dependent on man capability, there have been a lot of problems, such as losing a sale opportunity due to under stocking or having a management difficulty due to over stocking. To solve these problem. a joint reserch project entitled”Goods Oder Support Systems by Neural Networks” was started by local companies and KUT. This thesis reports a first-stage analysis results of selling tendency in POS data using self-correlation

*key words*     time series, Sample Auto Correlation

# 目次

|              |                                  |           |
|--------------|----------------------------------|-----------|
| <b>第 1 章</b> | <b>まえがき</b>                      | <b>1</b>  |
| <b>第 2 章</b> | <b>流通と情報化</b>                    | <b>2</b>  |
| 2.1          | 流通とは . . . . .                   | 2         |
| 2.2          | 流通の情報化 . . . . .                 | 2         |
| 2.2.1        | POS . . . . .                    | 3         |
| 2.2.2        | EOS . . . . .                    | 3         |
| 2.2.3        | VAN . . . . .                    | 4         |
| 2.2.4        | 小売業のネットワーク化 ～セブンイレブンの例 . . . . . | 4         |
| <b>第 3 章</b> | <b>発注予測と研究概要</b>                 | <b>6</b>  |
| 3.1          | 発注における変動要素 . . . . .             | 6         |
| 3.2          | 共同研究「自動発注支援システム」について . . . . .   | 7         |
| 3.3          | 予測情報の抽出プラン . . . . .             | 8         |
| <b>第 4 章</b> | <b>時系列解析と標本自己相関</b>              | <b>10</b> |
| 4.1          | 時系列解析の目的 . . . . .               | 10        |
| 4.2          | 標本自己相関関数 . . . . .               | 11        |
| <b>第 5 章</b> | <b>実験</b>                        | <b>14</b> |
| 5.1          | データ抽出の検討 . . . . .               | 14        |
| 5.2          | データ分析範囲 . . . . .                | 15        |
| 5.3          | 実験候補の抽出方法 . . . . .              | 16        |
| 5.4          | PERL プログラムによるデータ抽出 . . . . .     | 16        |
| <b>第 6 章</b> | <b>結果</b>                        | <b>21</b> |

## 目次

|              |                      |           |
|--------------|----------------------|-----------|
| 6.1          | IID ノイズ . . . . .    | 21        |
| 6.2          | 標本自己相関グラフ . . . . .  | 22        |
| 6.2.1        | 生理用品 A の結果 . . . . . | 22        |
| 6.2.2        | 洗髪料 B の結果 . . . . .  | 26        |
| 6.2.3        | 化粧品 C の結果 . . . . .  | 27        |
| <b>第 7 章</b> | <b>結果</b>            | <b>29</b> |
| 7.1          | 標本自己相関の結果 . . . . .  | 29        |
|              | <b>謝辞</b>            | <b>31</b> |
|              | <b>参考文献</b>          | <b>32</b> |

# 目次

|      |  |    |
|------|--|----|
| 2.1  | 流通循環図 . . . . .  | 3  |
| 3.1  | 共同研究図 . . . . .  | 7  |
| 3.2  | 予測情報の抽出アプローチ . . . . .                                 | 8  |
| 4.1  | 母集団からの標本抽出 . . . . .                                   | 11 |
| 5.1  | POS データの収録形式 (数値は仮のもの) . . . . .                       | 14 |
| 5.2  | PERL による抽出手順 . . . . .                                 | 17 |
| 5.3  | 化粧品 A 日別集計グラフ (1998 1/1~1998 12/31) . . . . .          | 18 |
| 5.4  | 化粧品 A 日別集計グラフ (1999 1/1~1999 12/31) . . . . .          | 19 |
| 5.5  | 化粧品 A 日別集計グラフ (2000 1/1~2000 10/31) . . . . .          | 19 |
| 5.6  | 化粧品 A 月別集計グラフ (1998 1/1~2000 10/31) . . . . .          | 20 |
| 6.1  | S-ACF グラフの構成 . . . . .                                 | 22 |
| 6.2  | 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$ : lag974 . . . . .  | 23 |
| 6.3  | 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$ : lag30 . . . . .   | 24 |
| 6.4  | 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$ : lag120 . . . . .  | 24 |
| 6.5  | 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$ : lag1~30 . . . . . | 25 |
| 6.6  | 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$ : lag50 . . . . .     | 26 |
| 6.7  | 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$ : lag100 . . . . .    | 26 |
| 6.8  | 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$ : lag150 . . . . .    | 27 |
| 6.9  | 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$ : lag974 . . . . .    | 27 |
| 6.10 | 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$ : lag30 . . . . .    | 27 |
| 6.11 | 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$ : lag60 . . . . .    | 28 |

## 図目次

|   |    |
|---|----|
| 6.12 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$ : lag90 . . . . .  | 28 |
| 6.13 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$ : lag974 . . . . . | 28 |

# 表目次

|     |                            |    |
|-----|----------------------------|----|
| 5.1 | 化粧品 A：日別集計結果（一部） . . . . . | 18 |
| 5.2 | 化粧品 A：月別集計結果（一部） . . . . . | 18 |
| 7.1 | 化粧品 A：売上ベスト 2 店舗 . . . . . | 29 |
| 7.2 | 化粧品 A：売上ベスト 3 店舗 . . . . . | 29 |

# 第 1 章

## まえがき

流通業者にとって商品の発注を適正に行うことは企業の競争力を左右する最大の課題である。適確な発注量によって商品在庫の円滑な循環が得られるが、これを読み誤れば在庫過少による販売機会の逸失や、在庫過大による経営の圧迫など企業の経営そのものに悪影響を与えかねない。今日のように消費者意識の多様化・複雑化が進んだ中で、需要に応じた商品を適切な数量だけ店頭に並べるのは発注担当者の経験則に依存するほかない。需要に応じた発注数量の最適化は、流通業における重要なキーファクターであり、メーカー、卸売業者、小売業者をつなぐ最適化ネットワークの実現に大きく寄与する所である。

本研究では変動的な要因や需要に対する発注作業の効率化をはかるため、自動発注支援システムを提案する。この研究は高知工科大学と高知県の流通企業、株式会社高知流通情報サービスおよび高知県の小売業者である株式会社サニーマートとの共同プロジェクトとして発起したものである。本稿は、このプロジェクトの初期研究の結果を報告するものである。まず第 1 章では、流通について簡単にまとめたものを記す。

## 第 2 章

# 流通と情報化

### 2.1 流通とは

我々人間は社会の中で生活するために様々なものを生産し、それを消費することで経済活動を行っている。ある商品が工場で生産され運送センターなどを経て店頭に並び、そしてその商品を消費者が購入して利用（消費）する。今日このような流れを繰り返して様々なものが流れている。しかしながら消費者のニーズは多様かつ変動的で、これに応じて流通業の供給面も変化しなければならない。ことに戦後の高度経済成長をむかえた日本では国民の社会生活自体が加速度的に変化し、情報化、国際化などと叫ばれるようになってからの消費者意識はますます複雑さを増していった。このことから流通業では情報技術によるシステム構築が活発化し、様々なシステムが組みられるようになった。今日では情報によって武装しなければこのような流動的なニーズには対応できない。人間の経済活動は様々なものを生産し、それを消費することで行われている。そのような中で流通業は、生み出された商品やサービスなどを消費者側に送るための重要な分野である。社会活動のおおまかな循環を図1に示す。

### 2.2 流通の情報化

コンピュータの発展とともに流通業においても1970～80年にかけて目覚ましいほどに情報技術の取入れが行われるようになった。流通業の情報化を説明する際によくPOS、EOS、VANが使われている。これらの技術の登場なくして流通の合理化はなかったと言われるほどである。情報管理システムを取り入れることによって、人的ミスの減少、迅速な情報管理、正確な処理など、これまで人手に頼っていた部分を補い、在庫、発注、売上管理の効率

## 2.2 流通の情報化

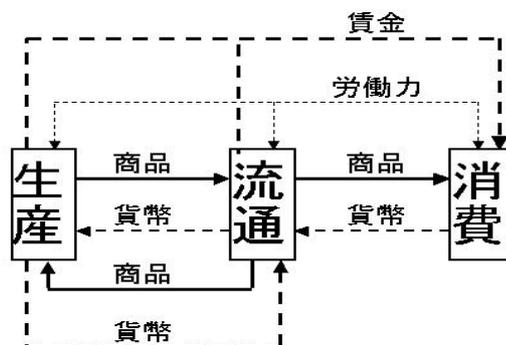


図 2.1 流通循環図

化を図るようになった。この章ではこの代表的な 3 つの技術を説明し、小売業情報化の例としてセブンイレブン社を挙げて説明する。

### 2.2.1 POS

POS(Point Of Sales : 販売時点情報管理) 商品データをバーコードなどから読み取り、単品情報、売上数量、日時などをデータベースに格納する。今日ではスーパーのレジスターなどに日常的に兼ね備えられている。このシステムを導入した結果、レジ作業の簡略化、省力化、タイプミス減少など店先での労働力の向上につながったほか、集計されたデータを利用して商品の売上推移や不良在庫の切り捨てなど、各部門の意思決定を判別する有効な情報として効果を発揮している。

### 2.2.2 EOS

EOS (Electronic Order System : 電子受発注システム) コンピュータと電話回線を利用して受発注を行うシステム。大手小売店やチェーン店ではすでに 100%に近い普及率を誇って

## 2.2 流通の情報化

いる.EOS を利用すればオンラインでの受発注が可能になるめ、情報交換の速度・精度ともに飛躍的に向上した.

### 2.2.3 VAN

VAN(Value Added Network : 付加価値通信網) 企業や異機種間のコンピュータを結びデータ交換を行うこと. 相手先に合わせてデータフォーマットの変換, メディア変換をするなど付加価値を加えて転送する技術のことである. すなわち, 流通のネットワーク化が進むにつれて増大した規格の相違を統一や変更するなどしてストレスなくオンライン上で情報をやり取りするために考案されたものである.

### 2.2.4 小売業のネットワーク化 ～セブンイレブンの例

上記に述べたような情報技術を取り入れ, 80年代の小売業では店舗ごとのネットワーク作りが盛んに行われるようになった. 組織的なネットワークの構築は流通業にとってこれまでにないほど効率的な成果をもたらすからである. 逆にいえばこれらの技術が登場して初めて流通業は流通業たるシステムの構築を行えるようになったのである.

店舗ごとのネットワークづくりが盛んになる中で, 最初にその成功例を示したのがセブンイレブン・ジャパンであった. セブンイレブンは1店舗に平均3000アイテムの商品を扱い,1日あたりの来店客数は全国で590万人にまでなる巨大なチェーン店である. セブンイレブンはこのPOSから得られる単品販売情報をもとに各店舗ごとの配送の計画化, 地域別集約化, 生産の連動などを独自に進め, 現在までに数度のシステム変更を行いながらこれを強化していった. これら物流における情報システムを導入した結果, 加盟店1店舗当たりの期末在庫額は半分以下となり, 加えて1日あたりの売上額は5割上昇という効果が得られたという.

今日の「適時, 適量, 適条件」という小売業の基本テーマを物流における主観としてとらえるようになったのもPOSが活躍しはじめてからであった. しかしながら, POSシステムは

## 2.2 流通の情報化

それ自体に有用性があるのではなく、あくまで販売したという過去のデータを保持しているのみである。収録された情報を適確に活用することで、初めて効果が得られるというものであることを留意しなければならない。その点を考慮した上でもセブンイレブンは成功者といえるのである。

## 第 3 章

# 発注予測と研究概要

前章までは流通の背景について大まかに述べた。本章からは具体的な研究目的とその流れを記す。

### 3.1 発注における変動要素

一般的な発注業務は現在においても人間の経験則から発生している。無論、先の章で述べたような POS など過去のデータベースからある程度の数量を計るものであるが、最終的な判断は長年現場で経験をつんだ発注担当者の能力に依存しているのである。それは各々の商品の売上要因となる変数値が膨大かつ複雑で、また発注商品は日々追加・更新されるためにシステムとしてとらえるには流動的すぎるからである。即ち、発注業務とは変動的な要因を過去の経験則を元に、適正な数量を予測する作業であると言い換えることができる。以下に売上の変動要因となるものをまとめてみた。

- 自然特性：季節、天候、気温など自然現象による売上の影響
- 地域特性：同一商品であっても町の中心部にある店舗と、山中や海沿いにある店舗などその地域固有の特性による売上の変動。商品の配送にかかる時間的な制約や循環効率は都心と地方ではばらつきが大きい
- 行事特性：スーパーなどのサービスまたはイベント（ポイントカードによる割引など）による売上変動や、他店がオープンした事による自社への影響、年末年始やシーズンによる変化など、特定の行事による売上率の変動

### 3.2 共同研究「自動発注支援システム」について

ここに挙げた例のほか、来店客層や店舗ごとの取扱商品の違いによる変動などがある。これまでも過去の販売データを元にした分析システムの構築が試みられたことがあったが、上記のように地域や商品による特性のばらつきが大きく標準的なシステムでの対応は困難とされてきたのである。そのため、在庫過小による販売機会の逸失や、在庫過大による経営の圧迫などといった問題を抱えている。本研究はこのような問題点を解決するため、変動特性の大きい発注業務をニューラルネットワーク等で算出し、小売業者のニーズに合った適正発注量を判断する商品発注支援システムの構築を目的とするものである。

### 3.2 共同研究「自動発注支援システム」について

上記のような背景のもとに本研究では変動的な在庫需要を予測し適正な発注量を判断するシステムの構築を立案した。図(3.1)のように流通企業である株式会社高知流通情報サービスおよび小売業者の株式会社サニーマートと、高知工科大学のラック研究室・竹田研究室が共同で進めているプロジェクト「ニューラルネットによる自動発注支援システムに関する研究」が本研究の背景である。

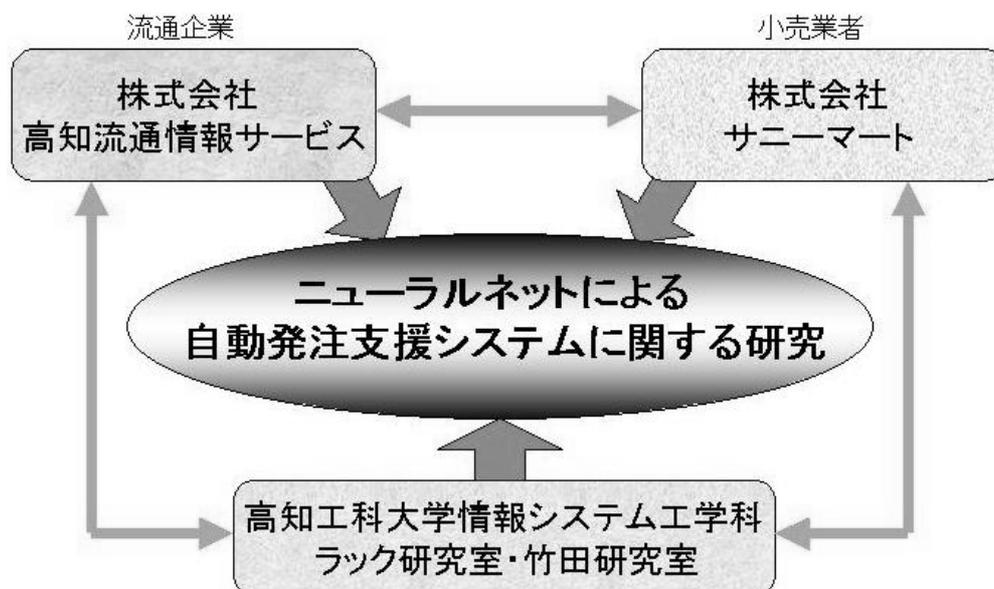


図 3.1 共同研究図

### 3.3 予測情報の抽出プラン

## 3.3 予測情報の抽出プラン

適確な発注量を予測するシステムには、その予測要因となる情報を抽出しなければならない。(4.1) にその予測情報の大まかな抽出アプローチを記す。

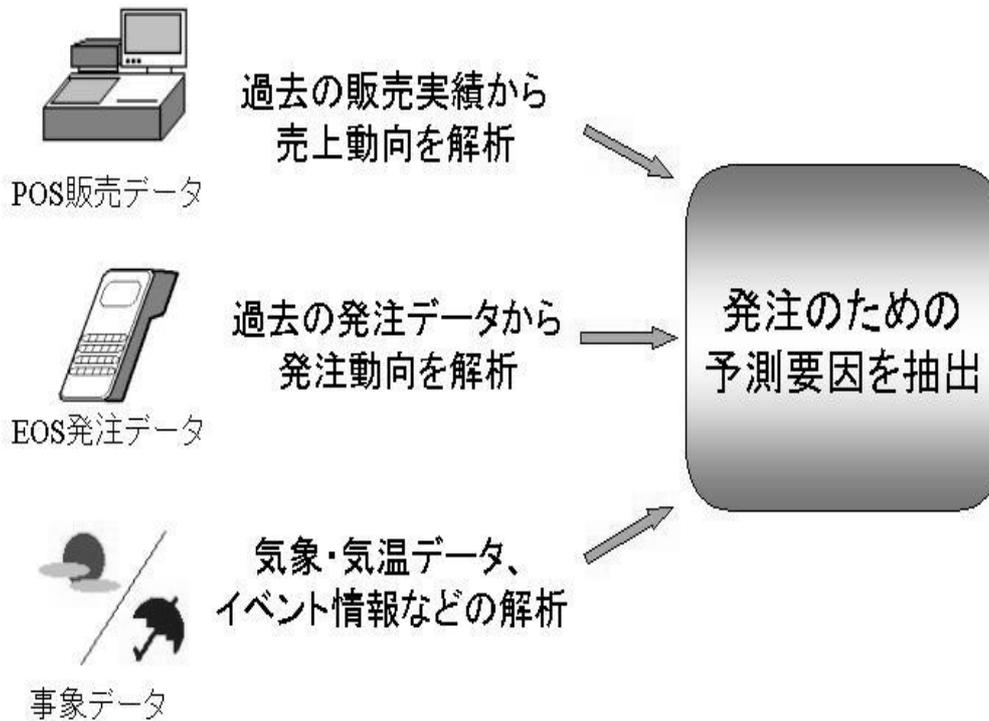


図 3.2 予測情報の抽出アプローチ

図 (4.1) は大まかなプランである。POS 販売データに収録された過去の商品販売の動向を調査することにより、売上傾向、顧客の購買傾向を調査する。このことにより将来の売上傾向を割り出し、予測情報として活用する。EOS 発注データからはこれまでの発注実績が記録されている。この情報を解析することにより商品の発注傾向を調査する。事象データからは、天候、気象、イベントなどの情報を調査し、商品の売上にどのように影響しているかを解析する。おそらく最も変動要因の大きいものだといえる。

上記のデータを吟味し、仮ルールによる仮想発注をシミュレートする。さらに過去のデータを元にした自己学習型（仮想発注→実際の販売データによる正解情報による学習）を繰り返す、ロジックの精度を向上させる。また、現場担に実際の操作を行ってもらい、そのノウハウ

### 3.3 予測情報の抽出プラン

を自己学習の正解情報として学習させることにより、地域や商品の特性に合わせたロジックの進化を起こそうとするものである。最終的に現場での活用を目的とし、消費者の多様なニーズに応えると共に、発注担当者の能力に依存せず均等な水準での発注を可能とする。

本研究は以上のような自動発注支援システムの構築を目的として立案された。本稿では、同プロジェクトの初期研究である POS 販売データからの売上動向からの情報をまとめると共に各商品の標本自己相関を調査するものである。

## 第 4 章

# 時系列解析と標本自己相関

### 4.1 時系列解析の目的

適正な発注量の決定をシステムで実現するには POS に収録された販売データを時系列によって解析し、これを元に予測モデルを立てる必要がある。すなわち、過去の販売動向を解析し、そこに何らかの傾向パターンが存在すれば、それを予測要因として抽出し活用できるという考えである。本章では時系列と標本自己相関関数について述べる。

時系列とはある時間軸上  $x_t$  に観測された観測値の集合のことであり各観測値は特定の時間  $t$  によって記録されているデータ集合のことである。POS データには膨大な商品が収録されているが、分析対象として有益なものは売上数値の高い商品となる。なぜなら、少ない標本症例にもとづいた結果は不確実性が高いと考えられるからである。

時系列解析において各観測値  $x_t$  即ち標本は、ある確率変数  $X_t$  の実現値と仮定する。これは、得られた時系列データの観測値は 1 つだけであるが、このデータを得るにあたり、その元となる母集団からたまたま観測された値であると考えるのである。統計的手法によって、ある対象の性質を調べる集団は非常に大きなデータから成るものである。例えば、日本人の平均体重を調査する場合、母集団とは全ての日本人ということになり、これらを全て調査するのは実質不可能である。つまり母集団全体の真の値は割り出すことはできない。そこで、おおよその推定値を得るために標本調査が必要となる。観測対象となる母集団からそれぞれのデー

## 4.2 標本自己相関関数

タ  $x_1, x_2, \dots, x_n$  を抽出し, その平均を割り出すことで母集団のおおよその平均値を求める. 母集団から一つの標本が観測される確率は, 母集団における相対頻度と考えることができ, また観測された標本値の平均は, 母集団平均の近似値を与えることになる. この「母集団における各値の相対頻度」を確率分布 (母集団分布) と考えると, 標本値は母集団分布という確率分布にしたがう確率変数となる. 抽出された標本値は他の標本値の観測に依存しないため, それぞれの標本は独立であるといえる. このような手法を統計的推測といい, より多くの標本を解析すれば, 母集団の性質を正確に理解できるという根拠につながっている.

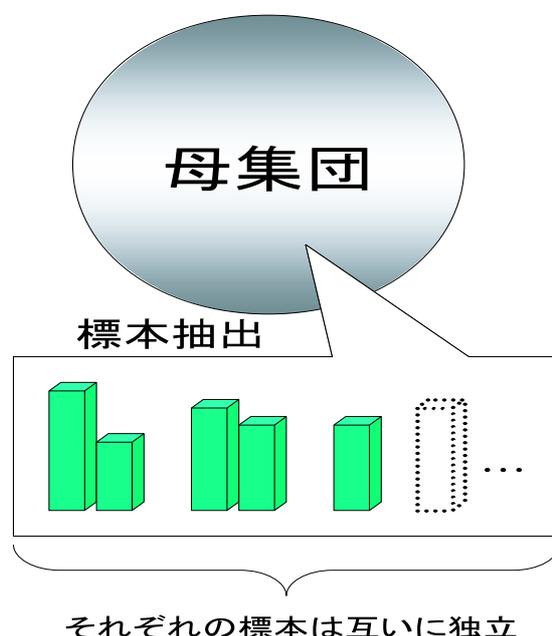


図 4.1 母集団からの標本抽出

## 4.2 標本自己相関関数

標本自己相関 (Sample AutoCorrelation Function : 以下 S-ACF) とは  $x_1, x_2, \dots, x_n$  をある時間軸上に観測された時系列データとした場合,  $x_t$  と  $h$  ずらしたデータ  $x_{t+h}$  に相関関係があるとき, これを標本自己相関という. S-ACF は以下のように定義されている.

$x_1, x_2, \dots, x_n$  をある時系列の観測値とする

## 4.2 標本自己相関関数

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad (4.1)$$

式 (4.1) を標本平均という。全観測値を加算した値を、要素数  $n$  で割っている。すなわち全体の平均値である。

$$\hat{\gamma}(h) := n^{-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) \quad (4.2)$$

式 (4.2) を標本共分散関数という。各観測値  $x_t$  とそれを  $h$  だけ、ずらした値の和の各項をそれぞれ標本平均で減じている。

ここではデータ数  $n - |h|$  ですべて加算しているが差分を  $n$  で除している。

$\hat{\gamma}(h)$  は、時間  $h$  離れたデータの結びつきの強さを表す。

式 (4.2) によって自己共分散が得られると、次式によって標本自己相関関数が求まる。

$$\hat{\rho} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad (4.3)$$

自己共分散によって求められた値を  $\hat{\rho}(0)$  で割る。これによって値を

## 4.2 標本自己相関関数

$$-1 \leq \gamma(0) \leq 1$$

と正規化している. なぜなら  $\hat{\rho}(0)$  はズレの無い自分自身を示すものであり, 今現在の自分との相関関係よりも強い相関というのはいない. すなわち

$$|\gamma(h)| \leq |\gamma(0)|$$

となるからである.

# 第 5 章

## 実験

### 5.1 データ抽出の検討

本研究は、サニーマトから提供していただいた POS 販売データを対象に時系列解析を行い、その売上動向に自己相関パターン分析を実施する。対象とした商品は、洗髪料、男性洗髪料、化粧品、生理用品、ペットフードの全 5 種類である。これらは長期的なスパンで見れば何らかの相関関係が見出せるのではという検討に基づいて選出された。

分析対象とした POS データは、全 5 種類は図 (5.1) のようにカンマ区切りの CSV 形式でテキストファイルに保存されている。

|     |   |        |   |   |   |      |   |               |   |            |   |        |   |        |   |        |
|-----|---|--------|---|---|---|------|---|---------------|---|------------|---|--------|---|--------|---|--------|
| 055 | , | 000102 | , | 1 | , | 0410 | , | 1231231231231 | , | メーカー ショウビン | , | 000001 | , | 000850 | , | 000850 |
| 026 | , | 000102 | , | 1 | , | 0410 | , | 4567890456789 | , | メーカー ショウビン | , | 000001 | , | 000728 | , | 000728 |
| 045 | , | 000102 | , | 1 | , | 0410 | , | 0123456789012 | , | メーカー ショウビン | , | 000001 | , | 000850 | , | 000850 |
| 039 | , | 000102 | , | 1 | , | 0410 | , | 9876543210987 | , | メーカー ショウビン | , | 000001 | , | 000850 | , | 000850 |
| 054 | , | 000103 | , | 1 | , | 0410 | , | 2468024680246 | , | メーカー ショウビン | , | 000001 | , | 000650 | , | 000650 |
| 045 | , | 000103 | , | 1 | , | 0410 | , | 1357913579135 | , | メーカー ショウビン | , | 000001 | , | 000650 | , | 000650 |
| 052 | , | 000103 | , | 1 | , | 0410 | , | 1122334455667 | , | メーカー ショウビン | , | 000001 | , | 000600 | , | 000600 |
| 045 | , | 000104 | , | 1 | , | 0410 | , | 1231231231231 | , | メーカー ショウビン | , | 000001 | , | 000600 | , | 000600 |
| 054 | , | 000104 | , | 1 | , | 0410 | , | 4567890456789 | , | メーカー ショウビン | , | 000001 | , | 000800 | , | 000800 |
| 026 | , | 000104 | , | 1 | , | 0410 | , | 0123456789012 | , | メーカー ショウビン | , | 000001 | , | 000698 | , | 000698 |

|       |     |    |      |        |     |      |    |    |
|-------|-----|----|------|--------|-----|------|----|----|
| 0     | 1   | 2  | 3    | 4      | 5   | 6    | 7  | 8  |
| 店舗コード | 年月日 | 曜日 | 分類番号 | JANコード | 商品名 | 売上数量 | 単価 | 売価 |

図 5.1 POS データの収録形式 (数値は仮のもの)

## 5.2 データ分析範囲

要素はそれぞれ

- 0：店舗コード（店舗識別のコード）
- 1：年月日（商品が実際に売れた日）
- 2：曜日（商品が売れた曜日）
- 3：分類番号（商品の種類を示す番号）
- 4：JANコード（商品を識別する番号）
- 5：商品名（商品の名称名）
- 6：数量（実際に売れた数量）
- 7：売価（商品の売上金額）
- 8：単価（その商品の単体の値段）

という構成になっている。ただしこれは分析のために用意された収録データであって、他店および一般の流通業においても実際に同じ形式で情報を管理しているわけではない。

## 5.2 データ分析範囲

商品の自己相関の解析にあたり、分析期間を1997年12月から2000年10月31日までの過去3年分のデータに絞り込むこととした。POSの単品情報は、実質10数年前から現在に至るまで蓄積されているが、一つの商品が10数年前から変わらず市場に存在するのは稀でしかないという問題が、共同研究の討議の結果、指摘されたからである。また、商品識別のためのJANコードは商品名などが変更されるたび（例：AシャンプーがNEW-Aシャンプーなど）に新しいJANコードに書き換えられ、同様の商品であっても識別コードが異なれば別の商品として捉えられてしまうのである。また、本研究の共同企業である小売業者のコメントによると「毎年次々と新製品が開発され市場に出回るが、どれも数年の間に姿を消す。しかし昔ながらの商品はいつまでも売れ、市場に残りつづけている」という。事例を挙げるとインスタントラーメン類やジュース類は毎年新製品が発売されるが、カップヌードルやコカコーラなど昔ながらの商品は現在も販売されているのである。このような事例のこと

### 5.3 実験候補の抽出方法

から、データ分析の期間を過去3年分に絞り込むこととした。

### 5.3 実験候補の抽出方法

全5種類のPOSデータから時系列解析に有用性のある情報を選出する。選出の対象となるのは販売実績の高い商品と、その商品を販売したすべての店舗である。データ抽出は独自に作成したPERLプログラムによって実施する。これは、提供されたPOSデータが1種類で数十メガを超えるテキストファイルであるため、Excelなどでの処理は容量的に不可能だからである。そこでプログラミング言語であるPERLを使用し、必要な情報抽出を実施する。PERLはテキストファイルやプロセスを簡潔に扱うために最適化されたインタプリタ言語であり、テキスト形式で収録されたデータを、効率的に抽出するのに適している。これを利用してPOSのデータファイルを処理する。

### 5.4 PERLプログラムによるデータ抽出

POSに収録された店舗コード・売上日時・商品識別データなどの情報をPERLプログラムによって抽出する。ここで販売実績の高い商品の選別と、店舗コード別にファイルを分割処理するなどした。具体的な選出手順は図(5.2)の通り。

同様の操作を全5品種について行う。この結果日別、月別にソーティングされた

## 5.4 PERL プログラムによるデータ抽出

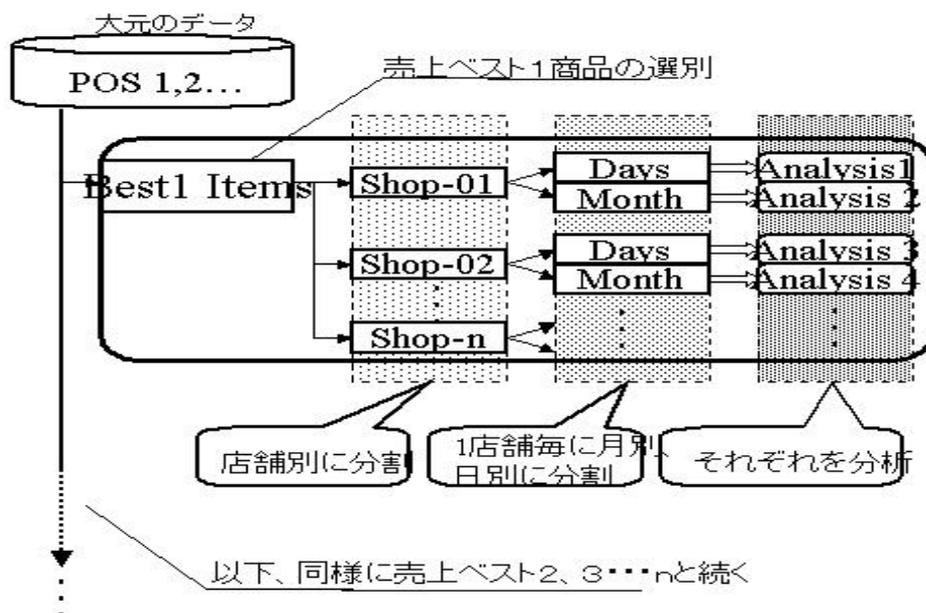


図 5.2 PERL による抽出手順

データが店舗コード別にファイルされる。店舗別への集計は、各店舗ごとの売上を比較するためであり、日別の売上集計は全3年間を通しての商品販売動向を見るものである。月別への集計はシーズンごとの変動を計るための処理である。最終的に、(7.1),(7.2)のように集計されたデータが得られる。

次に、整理されたデータファイルを Excel にてグラフ化する。ここでは横軸を時間として棒グラフを日別・月別共にプロットする。以下の図 (5.3)~図 (5.5) は化粧品 A の売上ベスト1店舗の日別売上データであり、図 (5.6) は同商品の月別グラフである。その他の分類に関しては、グラフ数量が膨大なため本稿では割愛する。

図中において 2000 年 9 月のデータ欠損は、店舗改装等の理由によるものである。以上のような手順で抽出・整理されたデータを用いて S-ACF を求める。次章にその結果を記述する。

## 5.4 PERL プログラムによるデータ抽出

表 5.1 化粧品 A : 日別集計結果 (一部)

| day\shop | shop2 | shop3 | shop4 | shop5 | shop6 |
|----------|-------|-------|-------|-------|-------|
| 980101   | 0     | 0     | 0     | 0     | 0     |
| 980102   | 0     | 0     | 0     | 0     | 0     |
| 980103   | 0     | 0     | 0     | 0     | 2     |
| 980104   | 3     | 4     | 2     | 0     | 1     |
| 980105   | 1     | 3     | 7     | 1     | 1     |
| 980106   | 2     | 3     | 2     | 0     | 0     |
| 980107   | 3     | 9     | 0     | 1     | 0     |
| 980108   | 2     | 3     | 0     | 1     | 1     |
| 980109   | 5     | 2     | 5     | 8     | 0     |
| 980110   | 0     | 2     | 3     | 0     | 2     |
| 980111   | 3     | 1     | 0     | 6     | 1     |
| 980112   | 0     | 2     | 2     | 0     | 0     |
| 980113   | 0     | 1     | 0     | 1     | 4     |
| 980114   | 0     | 1     | 2     | 2     | 0     |
| 980115   | 34    | 18    | 1     | 0     | 25    |
| 980116   | 6     | 0     | 0     | 0     | 0     |
| 980117   | 1     | 0     | 0     | 21    | 1     |
| 980118   | 2     | 1     | 0     | 1     | 0     |
| 980119   | 3     | 1     | 0     | 0     | 0     |

表 5.2 化粧品 A : 月別集計結果 (一部)

| day\shop | shop2 | shop3 | shop4 | shop5 | shop6 |
|----------|-------|-------|-------|-------|-------|
| Jan-98   | 91    | 198   | 87    | 80    | 74    |
| Feb-98   | 72    | 142   | 81    | 76    | 61    |
| Mar-98   | 71    | 82    | 62    | 58    | 57    |
| Apr-98   | 99    | 118   | 73    | 61    | 74    |
| May-98   | 153   | 151   | 82    | 99    | 26    |
| Jun-98   | 91    | 40    | 18    | 34    | 11    |
| Jul-98   | 102   | 74    | 60    | 30    | 38    |
| Aug-98   | 126   | 61    | 140   | 56    | 58    |
| Sep-98   | 75    | 57    | 32    | 43    | 92    |
| Oct-98   | 114   | 74    | 29    | 55    | 35    |
| Nov-98   | 88    | 26    | 35    | 34    | 77    |
| Dec-98   | 220   | 121   | 100   | 82    | 26    |
| Jan-99   | 72    | 38    | 43    | 18    | 24    |
| Feb-99   | 82    | 58    | 83    | 57    | 40    |
| Mar-99   | 112   | 96    | 177   | 74    | 74    |
| Apr-99   | 69    | 35    | 43    | 25    | 23    |
| May-99   | 54    | 77    | 31    | 6     | 76    |
| Jun-99   | 39    | 60    | 26    | 26    | 31    |

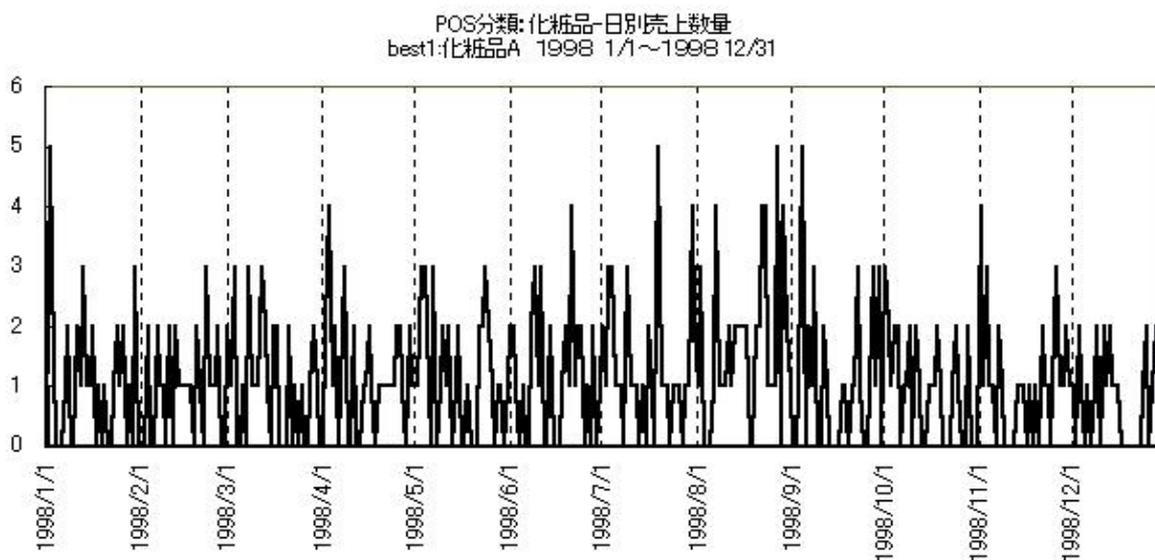


図 5.3 化粧品 A 日別集計グラフ (1998 1/1~1998 12/31)

## 5.4 PERL プログラムによるデータ抽出

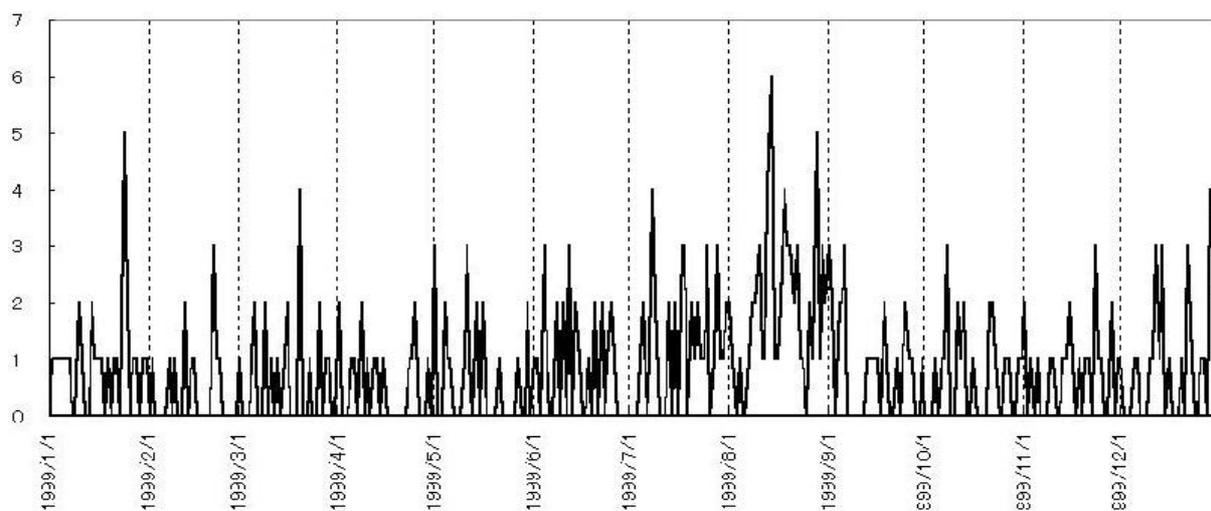


図 5.4 化粧品 A 日別集計グラフ (1999 1/1~1999 12/31)

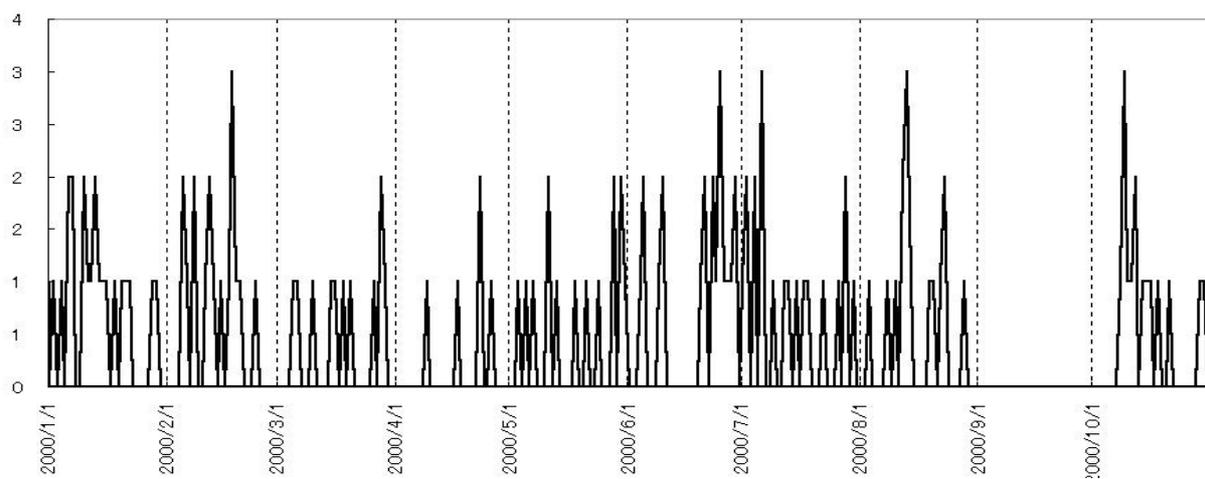


図 5.5 化粧品 A 日別集計グラフ (2000 1/1~2000 10/31)

## 5.4 PERL プログラムによるデータ抽出

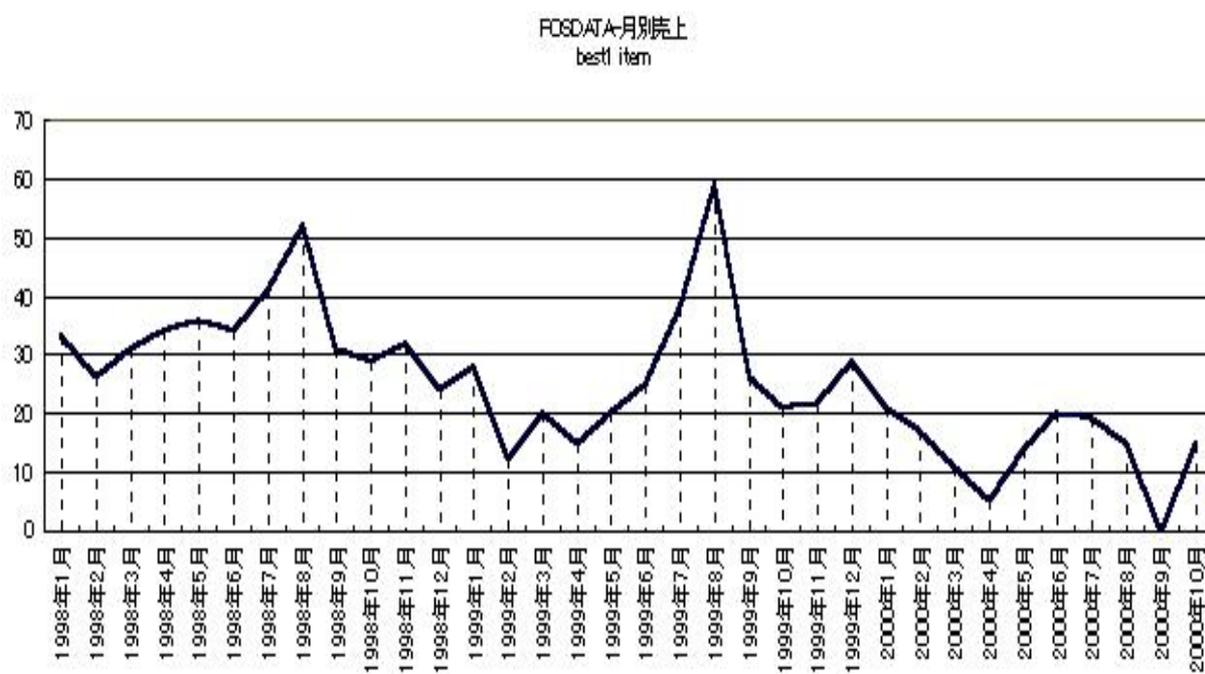


図 5.6 化粧品 A 月別集計グラフ (1998 1/1~2000 10/31)

## 第 6 章

# 結果

本章では MATLAB にてプロットされた S-ACF グラフを記述し、それぞれを検証する。まず、観測された標本値がどの程度の信頼度 (精度) を持っているかの区切りを明確にするため、IID ノイズを紹介する。

### 6.1 IID ノイズ

観測値が互いに独立同分布である平均 0 の確率変数の列  $X_1, X_2, \dots, X_n$  を IID ノイズ (independent identically distributed noise) という。これらは各観測値において、互いに依存しない独立した値であるため  $X_{n+h}$  の挙動を予測する場合には役に立たない。すなわちノイズとみなされる部分になる。

図 (6.1) は S-ACF グラフの構成図である。同図中における横破線を信頼限界区間という。信頼限界区間とは出力された値 (グラフの各々の縦棒) の 95% がこの区間内にとどまっていれば IID ノイズとみなされ、この区間を越えていけば自己相関の可能性があるとということを示している。

$n$  を全標本観測数としたとき、信頼限界区間は

$$\frac{\pm 1.96}{\sqrt{n}} \quad (6.1)$$

で得られる。ただし、1.96 は標準積分布の 0.975 確率点である。S-ACF は式 (4.3) によって  $|\gamma(h)| \leq |\gamma(0)|$  と規格化されているため、値が 1 に近いほど  $x_t$  と  $x_{t+h}$  の関係が強い

## 6.2 標本自己相関グラフ

事となる。即ち、互いに独立でない何らかの相関がある値が、この区間を超えるものと期待できる。

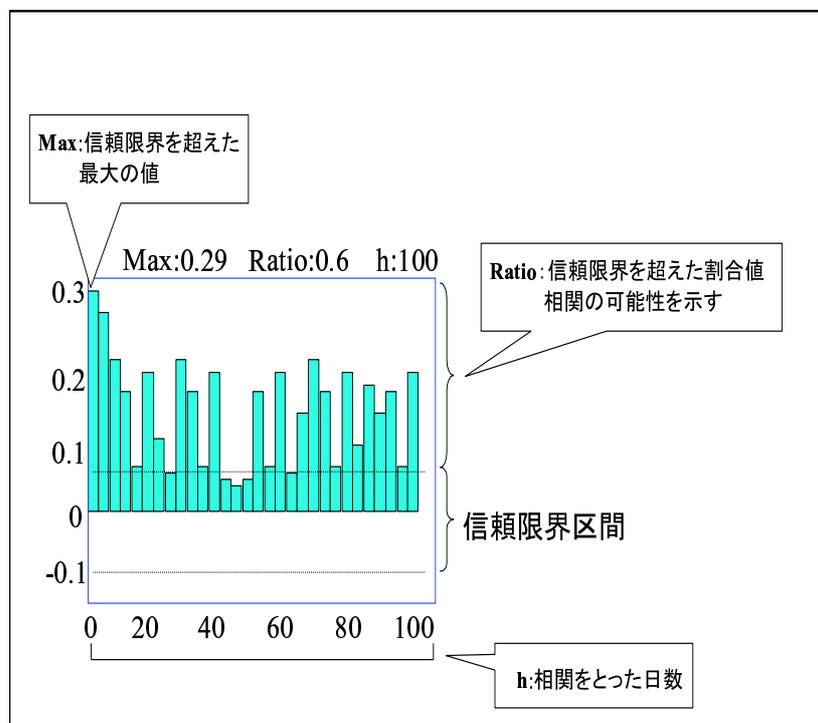


図 6.1 S-ACF グラフの構成

## 6.2 標本自己相関グラフ

以下に S-ACF グラフの出力結果を記す。ここで記載した相関グラフは特に相関の高かった「生理用品類, 化粧品類, 男性洗髪料類」の 3 つである。

### 6.2.1 生理用品 A の結果

図 (6.2) は生理用品 A, 売上ベスト 1 店舗の日別集計 S-ACF グラフである。グラフの横軸は lag であり、ずらした日数を表す。ここで lag974 は 3 年間に対して全ての日数をずらした結果ということになる。統計においてはこの高さ (MAX 値) が 0.6 を超える値が望ましいとされる。なお、この店舗  $\alpha$  は店舗改装などの不可抗力により一部のデータが欠損していたた

## 6.2 標本自己相関グラフ

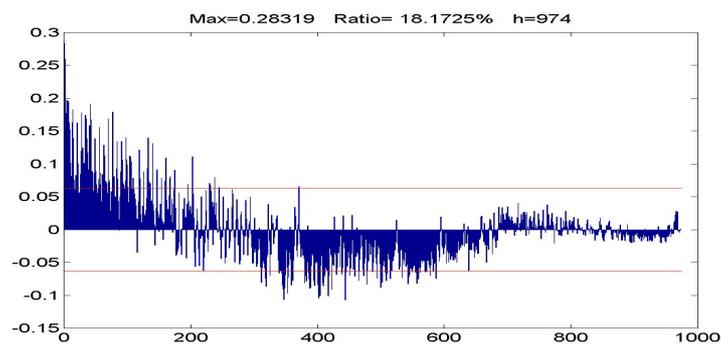


図 6.2 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$  : lag974

め, 集計期間を縮小して集計している.

図 (6.2) から, 信頼限界区間を超えている部分が, おおよそ lag0~lag200 に集中しているのがわかる. このずらし量 (lag) を変化させたグラフを以下に示す.lag の変化量は基本的に 30,60...120 と, 月単位でシフトさせることとした.

## 6.2 標本自己相関グラフ

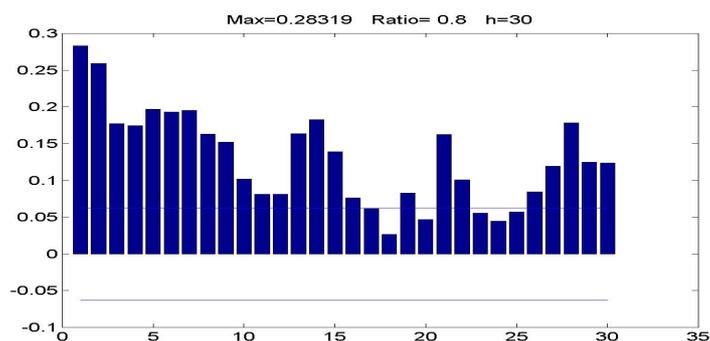


図 6.3 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$  : lag30

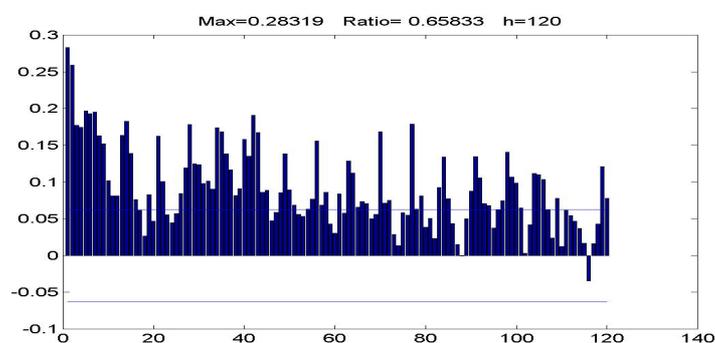


図 6.4 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$  : lag120

図 (6.2) のグラフを lag30 と lag120(すなわち 1 ヶ月分, 3 ヶ月分) と lag をずらし, 出力した結果が図 (6.3) と図 (6.4) である. これらの結果を見る限りでは相関の高さ (MAX 値) が有用的な数値ではないが, 割合の値 (Ratio) が両図ともに半分以上超える高さを示していることがわかる. また図 (6.4) には, 一定の間隔で相関が高くなっていることから, この商品に関して何らかの販売傾向があるとの見方ができる. また, 3 年間全体ではなく, ある部分のみを取り出して相関を取ることにする. 図 (6.5) がその結果である. これはある期間の相関のみを取り出すことによって余分なノイズが入らないようするためである.

## 6.2 標本自己相関グラフ

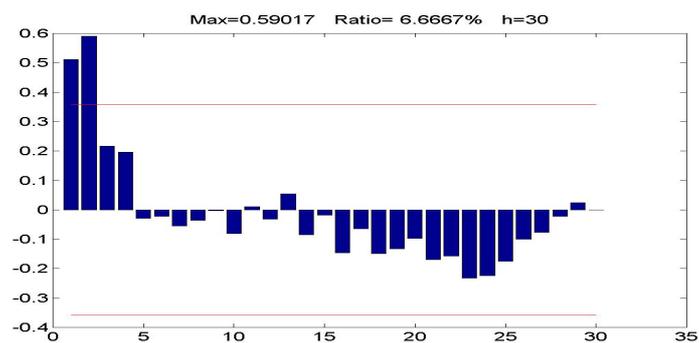


図 6.5 生理用品 A : 売上ベスト 1 店舗 shop- $\alpha$  : lag1~30

図 (6.5) のグラフは lag を 1 から 30 日間での間隔の自己相関結果である。これを見ると、lag1 と lag 2 に 0.6 に近い相関が得られている。このことは、今日の売上が明日・明後日の売上に何らかの影響を与えているということが言える。

## 6.2 標本自己相関グラフ

### 6.2.2 洗髪料 B の結果

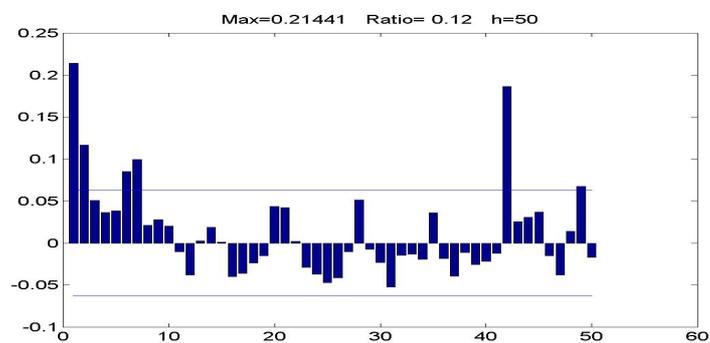


図 6.6 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$  : lag50

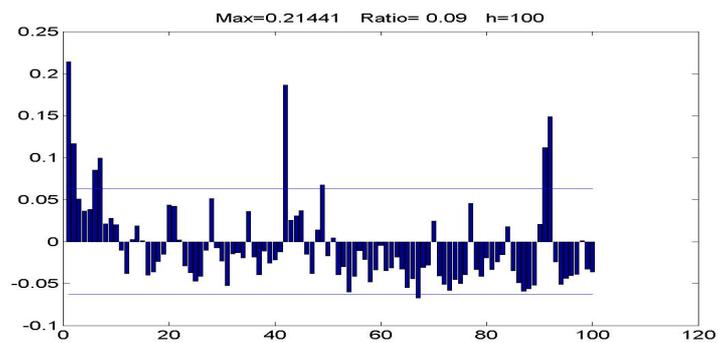


図 6.7 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$  : lag100

## 6.2 標本自己相関グラフ

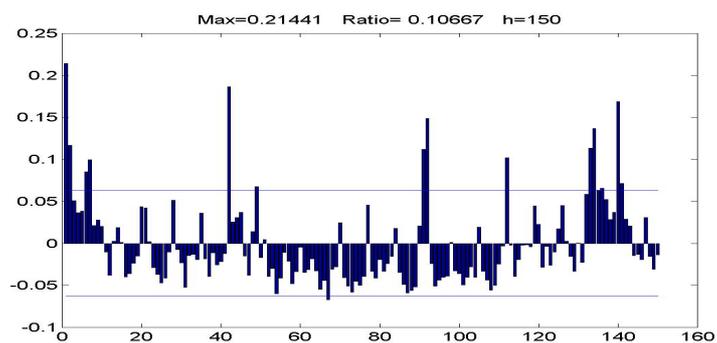


図 6.8 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$  : lag150

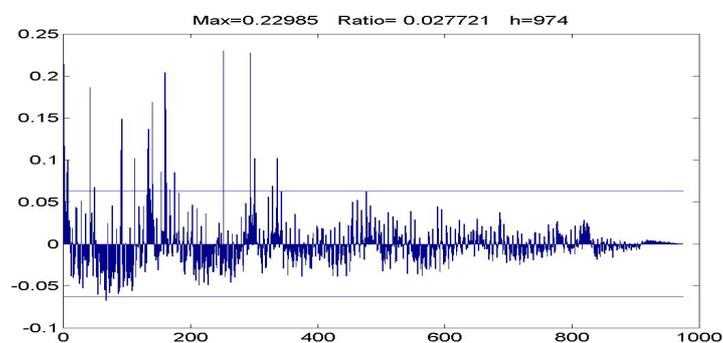


図 6.9 洗髪料 B : 売上ベスト 1 店舗 shop- $\beta$  : lag974

### 6.2.3 化粧品 C の結果

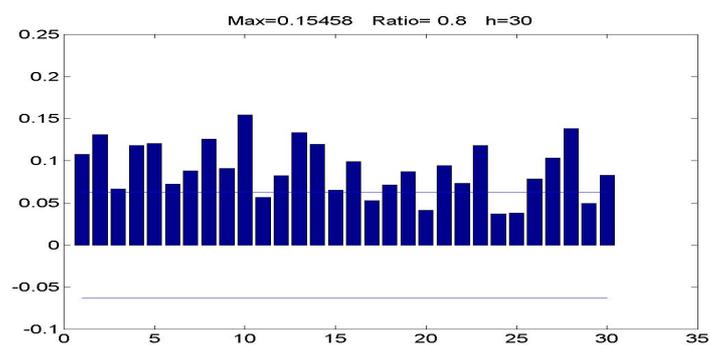


図 6.10 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$  : lag30

## 6.2 標本自己相関グラフ

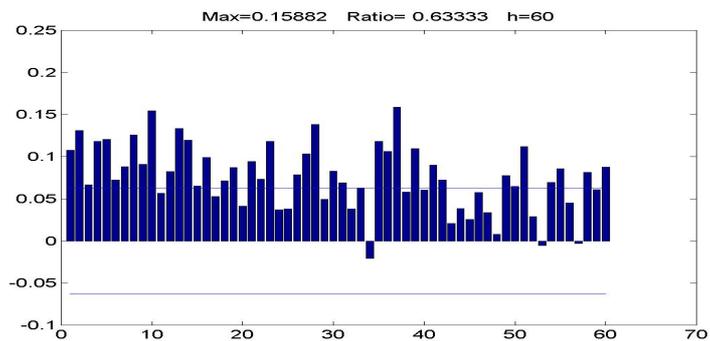


図 6.11 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$  : lag60

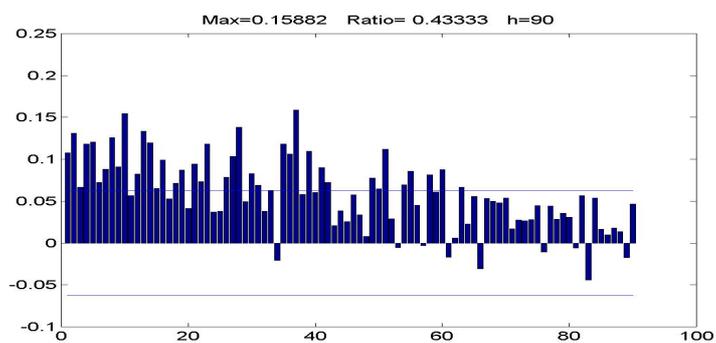


図 6.12 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$  : lag90

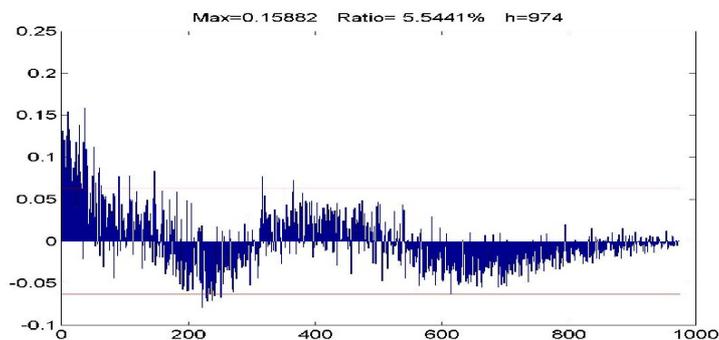


図 6.13 化粧品 C : 売上ベスト 1 店舗 shop- $\gamma$  : lag974

# 第7章

## 結果

### 7.1 標本自己相関の結果

提供された5種類のPOSデータに標本自己相関グラフをプロットした結果、生理用品類および化粧品類という女性が購入する商品について、信頼限界区間を越える割合値が得られた。とくに生理用品類は高い数値が得られたため、売上ベスト2店舗、ベスト3店舗と検証をした結果

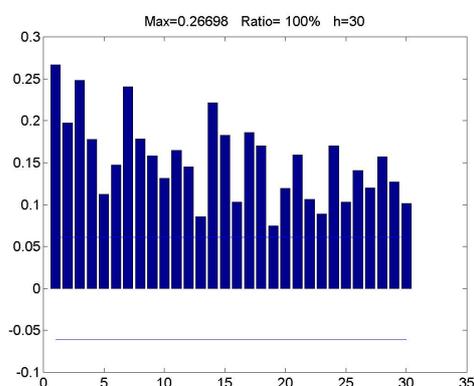


表 7.1 化粧品 A : 売上ベスト 2 店舗

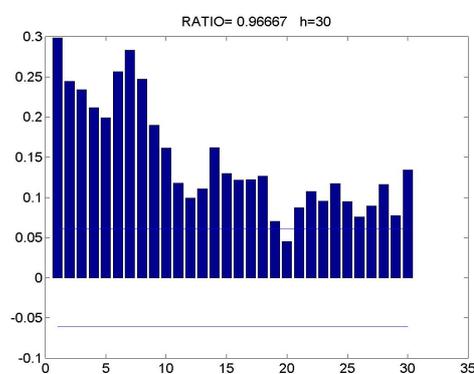


表 7.2 化粧品 A : 売上ベスト 3 店舗

といういずれも、MAX の値は高くないが、信頼限界区間を越えた割合が、ほぼ 100 % という高い割合が得られた。これらは何らかの購買傾向のパターンを示しているものと思われる。また男性洗髪料について、信頼限界を超える区間が約 1 週間間隔との傾向が現れた。これらの傾向は計算上に出された結果であるが、実際に現場でどのような影響を受けて、このような傾向を示しているのかというより複雑なデータの吟味が必要となる。

## 7.1 標本自己相関の結果

今回のデータ検証は一部の商品及び店舗を対象としたが、今後はより多種品目についてのデータ検証や、EOS 発注システムや事象データからの検証アプローチが求められる。

# 謝辞

本論分の作成にあたり, 有益なご指導をして頂いた Ruck Thawonmas 助教授に心より感謝を申し上げます. また, 大変貴重なアドバイスや情報を提供して下さった共同研究の株式会社 高知流通情報サービスおよび株式会社 サニーマートの皆様に心より御礼を申し上げます.

## 参考文献

- [1] P.J ブロックウェル / R.A デービス “入門 時系列解析と予測”(訳:逸見功),CAP 出版,pp1-25,July.2000.
- [2] 廣岡 治哉・野村 宏 “現代の流通” 成山堂出版,pp1-28,Aug.1994.
- [3] 野沢 健次 “現代流通入門” 清菱印刷,pp106-131,Nov.1997.