
ベイズの定理を用いた短期間購買データによる新商品需要予測モデルの提案

発表構成

- 研究背景・目的
- 既存モデル
- 実データへのあてはめ
- ベイズの定理
- まとめ・今後の方針
- 参考文献

東京理科大学大学院
工学研究科
綾部 英明

研究背景[7]

小売業界は、ダイエーの隆盛と期を同じくして「大量生産・大量消費」の時代であった

しかし近年、消費者のニーズの多様化が進み、消費者は様々な選択をするように変化している

そのため

- 各種メーカーにおいて、新商品の開発・投入の頻度が高まる
- 量販店はどの商品を売り場に出すかの判断が焦点となる

つまり

店舗側・メーカー側は
新商品の売れ筋・死に筋の
早期判断が求められている

と考えられる

研究背景

- ▶ 「商品の寿命は3週間」・「多産多死は『ショートセラー』で勝つ」
[9]とも言われている
- ▶ 新商品の需要予測を行うことでメーカーは製販調整・チャネルの拡大・縮小などの意思決定を早期に行うことができる
- ▶ しかし、市場導入直後の短期間での購買データを利用した新商品の需要予測モデルの研究はほとんど行われていない
[10]

※本研究では、「新商品」を既存の商品カテゴリーにおいて新たに参入する商品ブランドと定義する。
(Ex. お茶飲料カテゴリーにおける新ブランド『伊右衛門』・『生茶』など)

研究目的

新商品の需要を短期間の購買データから予測するモデルを提案する

(※本研究では「短期間」を2~3週間と捉える)

- 参照する既存モデルとして「社会的なブームの微分方程式モデル[1]」を利用する
 - ✓ 既存モデルは、過去の事例に関して当てはめを行っただけに過ぎない
- “予測”には、「ベイズの定理」を用いることとする
 - ✓ 過去のデータを利用することで、自商品のデータが少量でも活かす事が可能

既存モデルの模式図

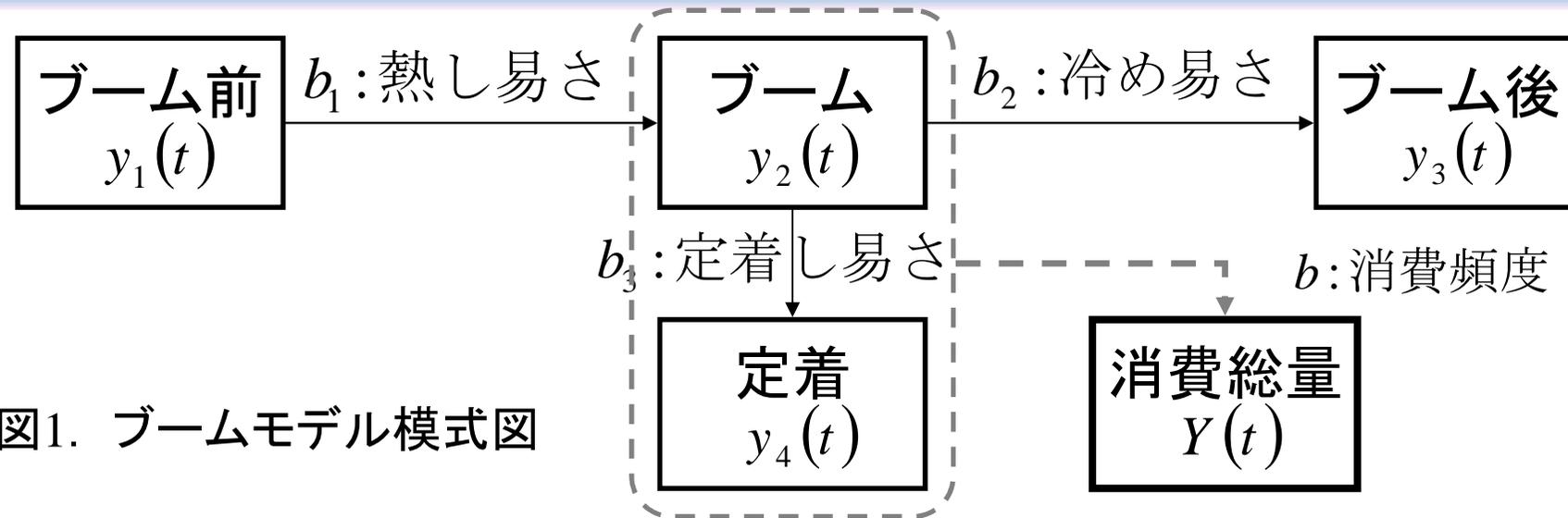


図1. ブームモデル模式図

$y_1(t), y_2(t), y_3(t), y_4(t)$: 時刻 t における顧客数

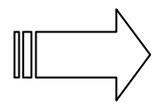
b_1 : 単位時間内にブームに乗じて製品消費を開始する未消費者変化率

b_2 : 単位時間内に製品に飽きて製品消費を中止するブーム消費者変化率

b_3 : 単位時間内に製品を気に入り定着層に移行するブーム消費者変化率

顧客の状態と定式化

- ▶ 《ブーム前》: 新商品登場前, 考察対象商品を消費していない状態
- ▶ 《ブーム》 : トライアル購買を行って商品を消費している状態
- ▶ 《ブーム後》: 商品に不満を抱き消費をやめた状態
- ▶ 《定着》 : 単位時間に一定割合で商品を消費し続ける状態



本研究では, 既存モデルで指す「ブーム」という現象を
トライアル購買によるものとする

- ブームが始まってからの状態変化
 - ✓ 《ブーム前→ブームへの変化》
 - ✓ 《ブーム→ブーム後への変化》
 - ✓ 《ブーム→定着への変化》

$$y_1'(t) = -b_1 y_1(t) \quad \dots(1)$$

$$y_2'(t) = b_1 y_1(t) - (b_2 + b_3) y_2(t) \quad \dots(2)$$

$$y_3'(t) = b_2 y_2(t) \quad \dots(3)$$

$$y_4'(t) = b_3 y_2(t) \quad \dots(4)$$

消費・非消費顧客数

- ▶ 時刻 t までの累積消費量

$$Y(t) = \int_0^t b \{y_2(t) + y_4(t)\} dt \quad \dots(5)$$

b : 消費頻度パラメータ[消費単位/(人・時間)]

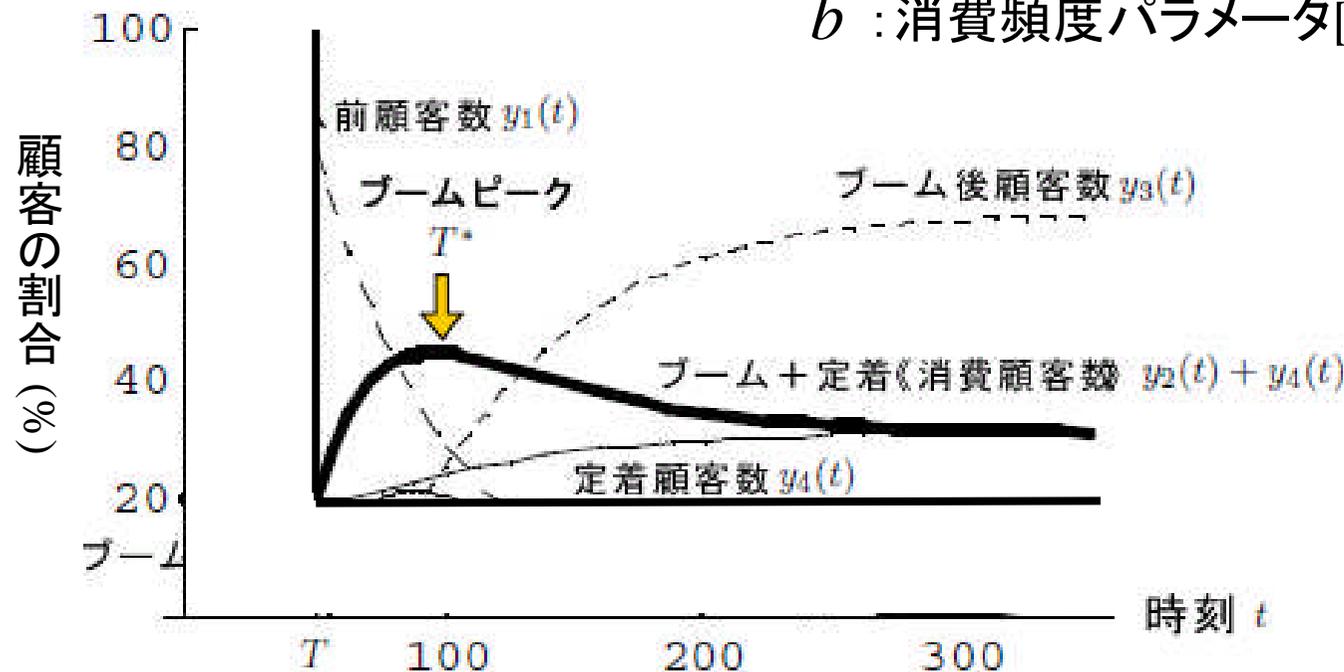


図2. ブームモデルの解例 $(S, k, T, b_1, b_2, b_3) = (100, 0.2, 50, 0.02, 0.03, 0.005)$

実データへのあてはめ

既存モデルは、
一般的に呼ばれる様々な“ブーム”という現象を捉えたものである
(Ex. 即席めんブーム・サッカー観戦ブーム・特定の単語が新聞記事に登場する回数の推移)

商品の中でも一つの新商品という小さなカテゴリーにおいて
当てはめを行い、モデルの妥当性を検証することとする

- ▶ データ収集期間の間に新発売された「お茶」についてあてはめを行う
 - ✓ 2005年3月に発売された日本コカ・コーラ社の「一(はじめ)」を対象とする
 - ✓ 一番データの多いPETボトル500mlを対象とする(616件)



既存モデル

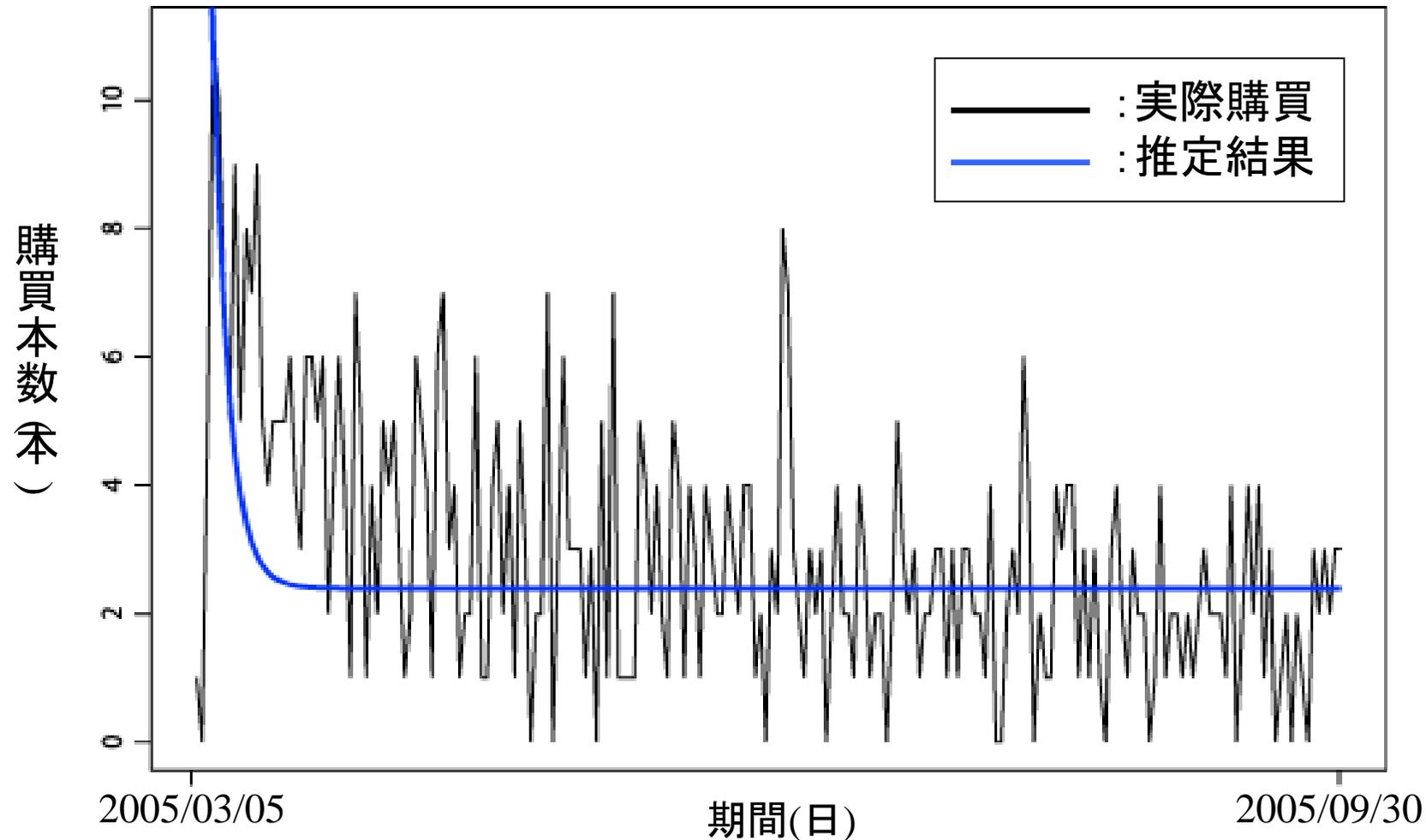


図3. 実際購買と推定結果(既存モデル)

既存モデル

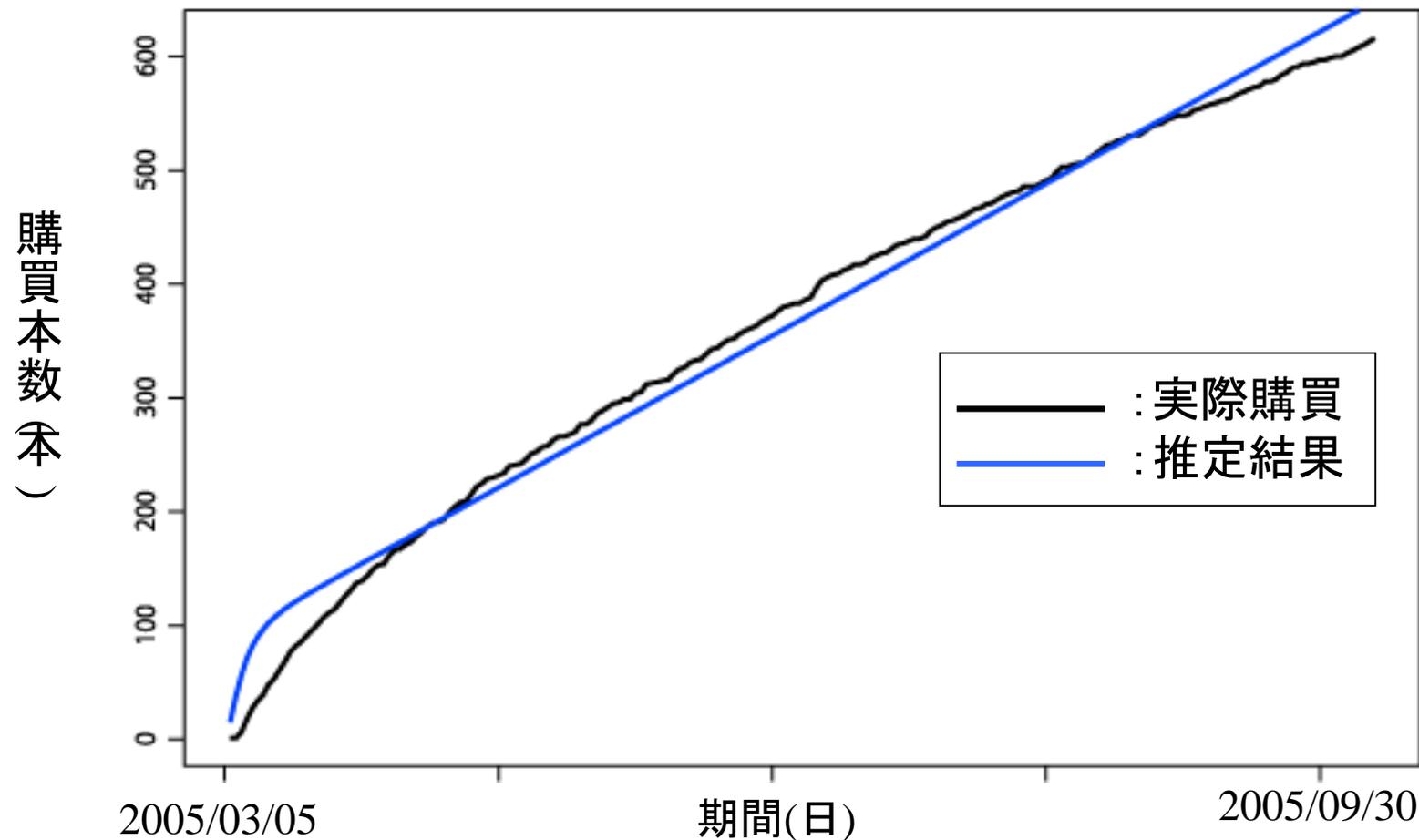


図4. 実際購買と推定結果(既存モデル)

ワイブル分布のモデル

- ▶ 売上分布の推定には、指数分布よりもワイブル分布の当てはまりがよい[5]
 - ✓ 既存モデルと比較を行うため、ワイブル分布において推定を行うこととする

ワイブル分布を組んだモデル

$$Y_t = m \left(1 - e^{-\left(\frac{t}{\phi}\right)^\gamma} \right) \dots (6)$$

t : 期間

Y_t : 時刻 t における累積販売数

m : 実際の総購買本数

ϕ, γ : パラメータ

ワイブル分布のモデル

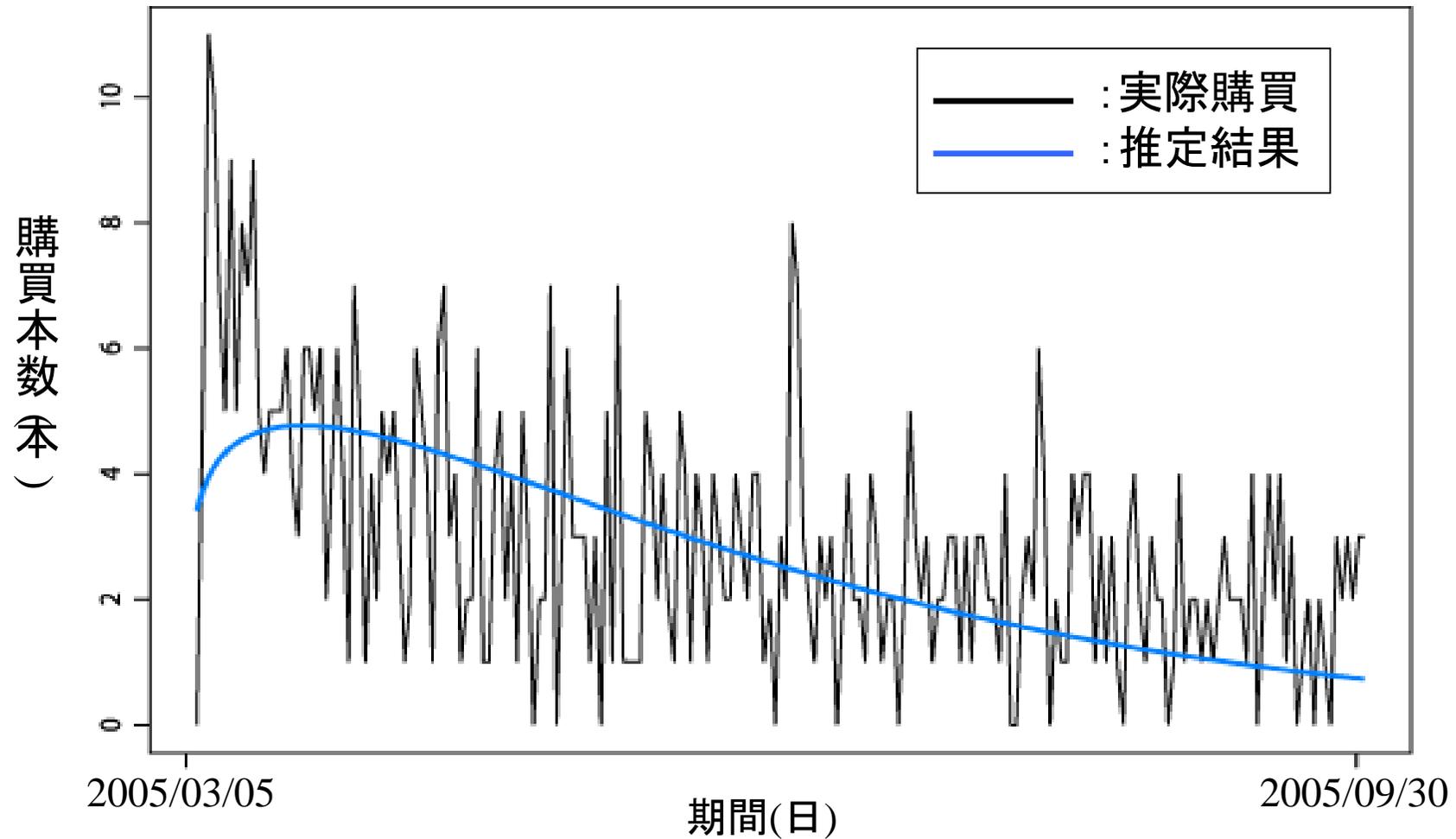


図5. 実際購買と推定結果(ワイブル)

ワイブル分布のモデル

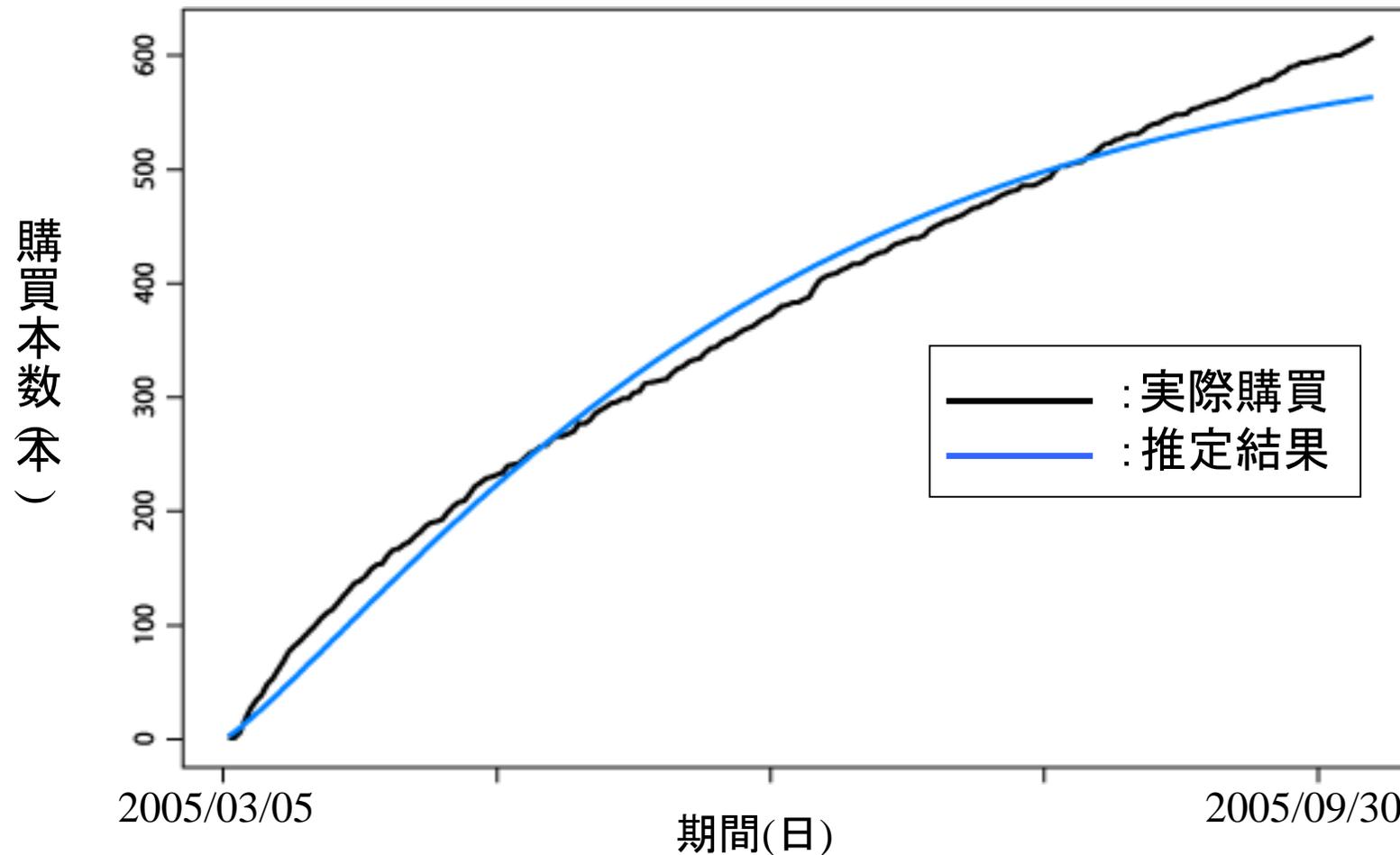


図6. 実際購買と推定結果(ワイブル)

指数モデル

- ▶ 既存モデル・ワイブル分布のモデルのように算式が複雑なもの（パラメータの数が多い等）を用いる必要があるのかを検証するため基本的な指数モデルでもあてはめを行う

$$\left\{ \begin{array}{l} y_t = e^{(\beta_0 + \beta_1 t)} \quad \dots(7) \\ \log y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad \dots(8) \end{array} \right.$$

β_0, β_1 : パラメータ

ε_t : 誤差

指数モデル

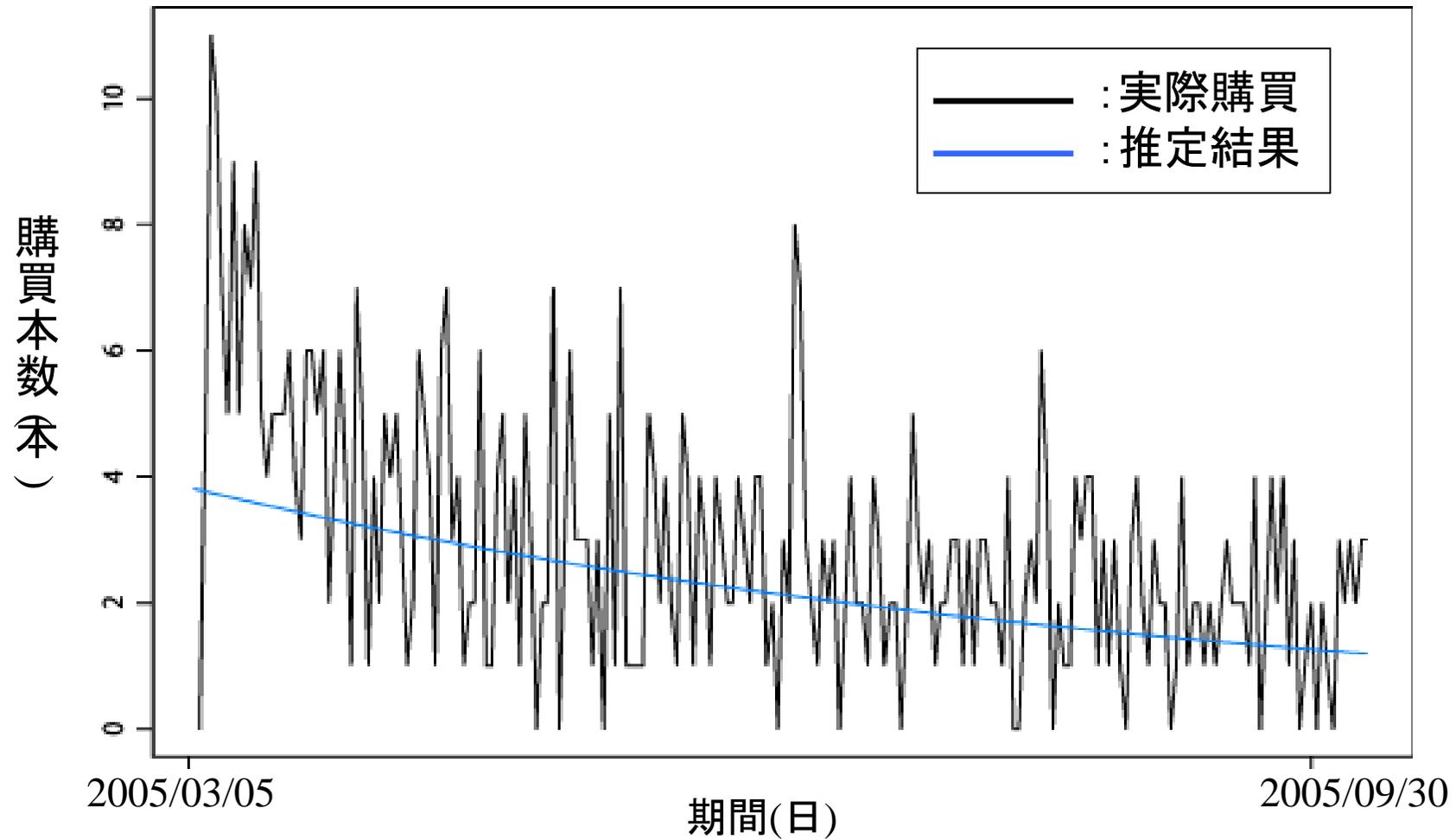


図7. 実際購買と推定結果(指数)

各モデルのあてはまり

- ▶ 各モデルのあてはまりを比較するためAIC (Akaike's Information Criterion)を求めた

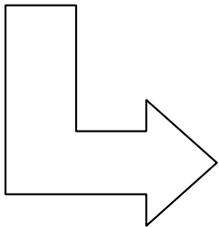
表1. 算出した各モデルにおけるAIC

モデル	AIC
既存モデル	-1436.06
ワイブル分布モデル	-1415.80
指数モデル	601.23

- AICの値からも、単純な指数モデルと既存モデル・ワイブル分布モデルの当てはまりのよさが伺える
- 既存モデルとワイブル分布モデルの当てはまりは同等と見受けられ、既存モデルを利用することの妥当性はあると判断できる
- 既存モデルは、ワイブル分布モデルよりパラメータの数は二つ多いが、最終的に0に収束しないことも考えられるという利点がある

既存モデルの問題点と解決法

- ▶ 既存モデルは、過去の事例に関して当てはめを行い、モデルの妥当性を述べたに過ぎない
 - ✓ すなわち、予測を行うことを考えて作られたものではない



ベイズの定理を利用する

- 自商品の短期間の販売データから得られる情報と過去の類似商品の販売データから得られる情報を利用して、自商品の未だ得られていない期間(将来)の推測が行える
- 推測が行った後のデータは、得られた情報としてその後に利用することができる(更新が可能)

ベイズの定理

- ▶ ベイズの定理は(9)式のように表せる

$$p(b, b_1, b_2, b_3 | y) \propto l(b, b_1, b_2, b_3 | y) p(b, b_1, b_2, b_3) \quad \dots(9)$$

$p(b, b_1, b_2, b_3)$: 未知パラメータ b, b_1, b_2, b_3 についての事前の確率密度(事前分布)

$p(b, b_1, b_2, b_3 | y)$: $Y_t=y$ が与えられたときの b, b_1, b_2, b_3 の事後の確率密度(事後分布)

(※ y : 新商品の得られた短期間の販売データ)

$l(b, b_1, b_2, b_3 | y)$: b, b_1, b_2, b_3 についての尤度関数(尤度)

ベイズの定理を用いた本研究の予測フロー

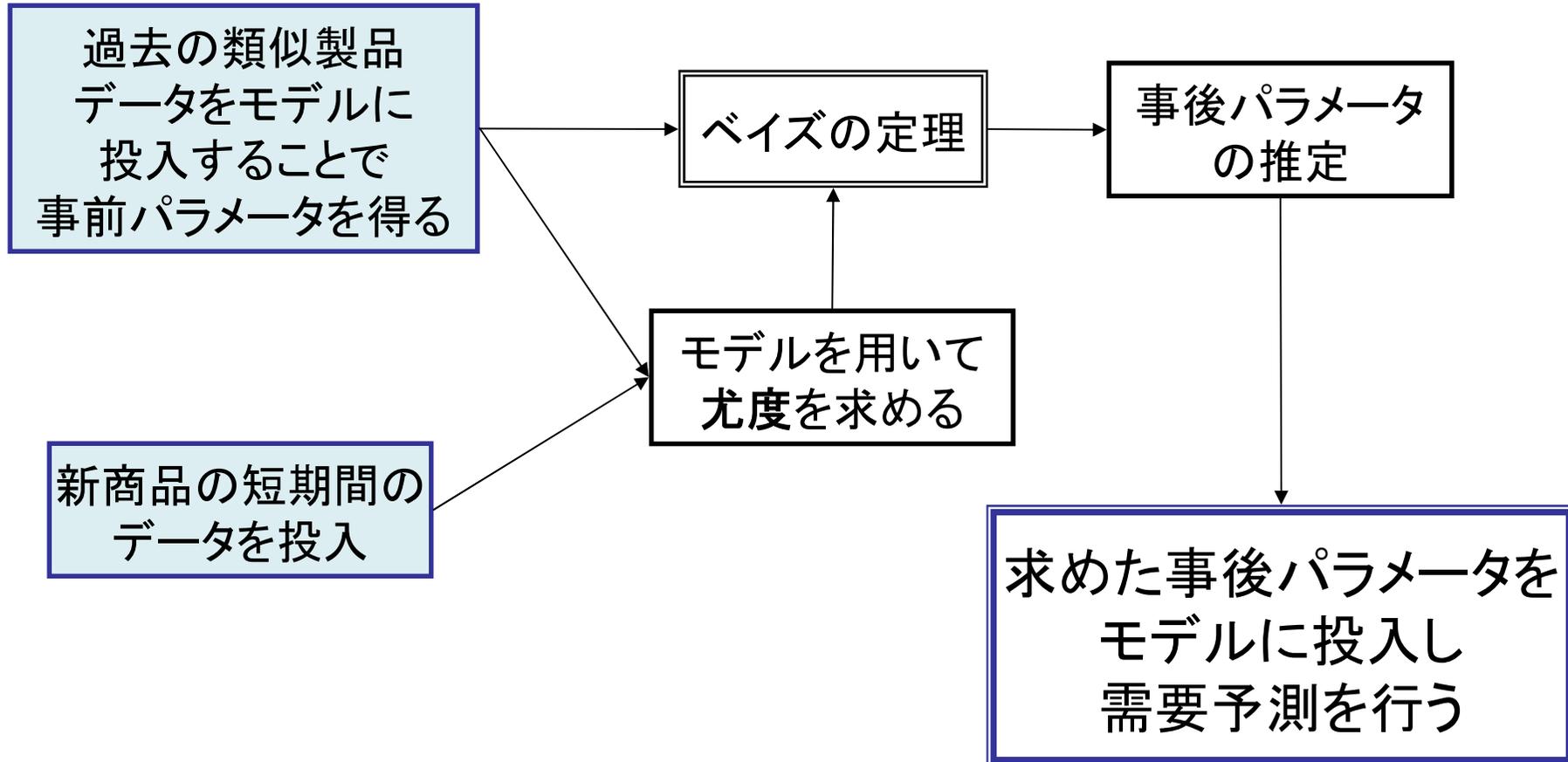


図8. ベイズを用いた本研究の予測フロー

尤度の計算

▶ 尤度の計算は以下の式にて算出する

$$\hat{Y}(t) = f(b, b_1, b_2, b_3) + \varepsilon_t \quad \dots(10)$$

$$Y(t) - \hat{Y}(t) = \varepsilon_t \quad \dots(11)$$

$$\varepsilon_t = Y(t) - f(b, b_1, b_2, b_3)$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

$$l(\varepsilon_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right) \quad \dots(12)$$

$\hat{Y}(t)$: 推測された購買本数

$Y(t)$: 実際の購買本数

$f(b, b_1, b_2, b_3)$

: パラメータ b, b_1, b_2, b_3 の関数

ε_t : 予測誤差

$l(\varepsilon_t)$: 尤度関数

まとめ・今後の課題

- ▶ マクロ的な“ブーム”に関するモデルに対し、ある一つの新商品の事例を当てはめても、モデルとして妥当であると思われる
- ▶ ベイズの定理を利用して需要予測に繋げるフローをまとめた
- ▶ 事前パラメータ・尤度は、正規分布を仮定し、事後パラメータの推測を行う
- ▶ 実データで実証分析を行い妥当性を検証する
 - ✓ ただし、正規分布の仮定は要検討する必要がある

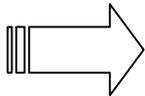
参考文献

- [1]中桐裕子, 栗田治 “社会的なブームの微分方程式モデル”, 日本オペレーションズ・リサーチ学会和文論文誌 Vol.47 pp.83~105 (2004)
- [2]フィリップ・コトラー, ゲイリー・アームストロング:
「コトラーのマーケティング入門(第4版)」, トッパン(1999)
- [3]フィリップ・コトラー著, 恩蔵直人監修, 月谷真紀訳:「コトラーのマーケティングマネジメント
ミレニアム版」, ピアソンエデュケーション(2001)
- [4]蓑谷千凰彦著, 「統計分布ハンドブック, 朝倉書店(2003)
- [5]Masataka Yamada, Hiroshi Kato: “A Structural Analysis of Sales Patterns of Music CDs”,
under revision(2002)
- [6]山田正孝, 加藤宏:「音楽CDの売上パターンの構造的分析」, 日本マーケティング・サイエンス学会(2002)
- [7]里村卓也, 森雅夫:「消費者の新製品受容過程とブランド選択
～試行段階のある新製品の反復購買モデル～」, マーケティング・サイエンスVol.8, No1・2
pp15-29 (2000)
- [8] 渡部洋著, 「ベイズ統計学入門」, 福村出版(2004)
- [9]「商品の寿命は3週間」, 日経ビジネス2006. 5.29号 pp26~29
- [10]中村博:「新製品のマーケティング」, 中央経済社(2001)

Appendix

S-PLUS・S+NUOPTでのプログラム

- ▶ 既存モデルのパラメータの推定・実データの当てはめ
- ▶ ワイブル分布のモデルのパラメータの推定・実データの当てはめ
- ▶ 指数モデルのパラメータの推定・実データの当てはめ

 別紙(Microsoft Word形式 「S-PLUS学生研究奨励賞.doc」)
に添付してあります。

Appendix データ概要

➤ 内容

- ✓ お茶に関するPOSデータ
(株式会社インテージ
personal eye academic版より提供)

POS (Point of Sales)データとは
販売時点管理データ, つまり店舗ごとに
「いつ」「何を」「いくつ」購買したのかを
把握することができる

➤ 期間

- ✓ 2004年10月1日～2005年9月30日の1年分

➤ データ内容

大きく分けて購買履歴データと購買者データに大別される

✓ 購買履歴データ:

購入日, 購入時間, JANコード, 商品名, 購入数,
内容量, 製造メーカー, 購入ルート, 詳細ルート

✓ 購買者データ:

モニター番号, 性別, 年代, 未既婚, 職業, 収入, 住居エリア,
勤務地エリア, インターネット利用の有無, 家族人数, 子供人数, 末子年齢,
家族構成, 車の有無, 生協加入の有無

➤ データレコード件数

- ✓ 購買履歴件数: 19873件 (503名)

Appendix データ例

表3. 使用したデータ(抜粋)

モニター	購入日	時間	JANコー	JANSEQ	JAN名称	個数	容量	単容量	単容量3	AM002	メーカー	AM006
1	20050207	1457	4901085096	0		1	500	500	500	71		5
1	20041031	1057	4901085080	0		1	500	500	500	71		5
1	20050304	1953	4901085071	0		1	500	500	500	71		5
1	20041109	2014	4901085096	0		1	500	500	500	71		5

容器	AM008	入り数	AM201	標準ルー	AM202	詳細ルー	AM232	使用者	AB304	性別	AB306	年齢②
ボトル	1	1本	1	コンビニエ	1		1	本人使用	1	男	3	30-39歳
ボトル	1	1本	2	スーパー	19	その他スー	2	共用	1	男	3	30-39歳
ボトル	1	1本	1	コンビニエ	1		1	本人使用	1	男	3	30-39歳
ボトル	1	1本	2	スーパー	19	その他スー	1	本人使用	1	男	3	30-39歳

AB307	未既婚	AB308	職業	AB310	個人収入	AB311	インター	AB312	プレイス	AB313	勤務地
2	未婚	4	派遣社員・契	4	200~299	2	利用してい	1	新宿エリア	12	その他東京(
2	未婚	4	派遣社員・契	4	200~299	2	利用してい	1	新宿エリア	12	その他東京(
2	未婚	4	派遣社員・契	4	200~299	2	利用してい	1	新宿エリア	12	その他東京(
2	未婚	4	派遣社員・契	4	200~299	2	利用してい	1	新宿エリア	12	その他東京(

AB318	家族人数	AB321	子供人数	AB322	末子年齢	AB323	家族構成	AB325	車	AB326	生協加入
1	1人	1	子供なし	1	子供なし	1	単身	1	なし	2	加入していない
1	1人	1	子供なし	1	子供なし	1	単身	1	なし	2	加入していない
1	1人	1	子供なし	1	子供なし	1	単身	1	なし	2	加入していない
1	1人	1	子供なし	1	子供なし	1	単身	1	なし	2	加入していない

- モニターの募集方法： 調査員リクルート・公募
- 調査方法： 購入商品のバーコードスキャニング方式

Appendix[10] 新商品の販売推移パターン

- ▶ CVS等では，導入直後の売上が最も高くなる傾向にある
 → 需要予測のモデル式は，導入時の売上推移パターンと長期の減衰パターンにより規定されると考えられる

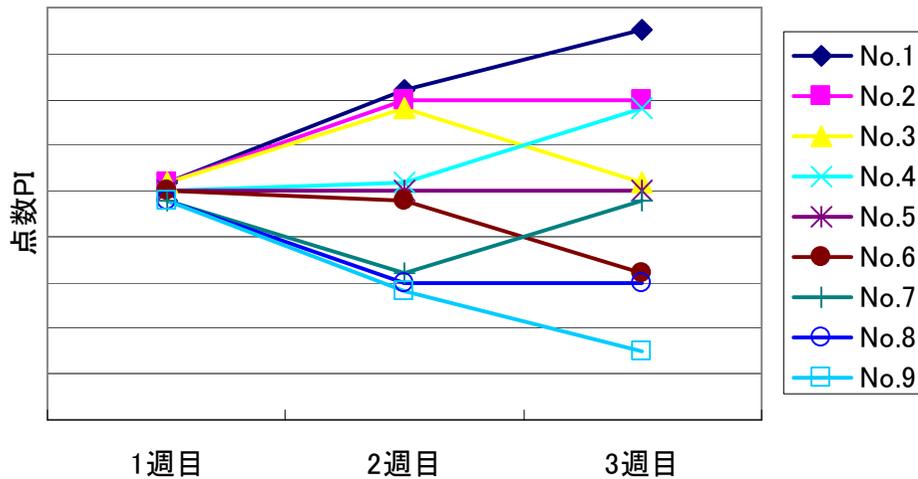


図9. 推移パターン

表2. 推移パターン頻度分布

	推移パターン		アイテム数	構成比
	1週目～2週目	2週目～3週目		
No.1	上昇	上昇	62	3.6%
No.2	上昇	横ばい	52	3.0%
No.3	上昇	下降	239	13.8%
No.4	横ばい	上昇	26	1.5%
No.5	横ばい	横ばい	33	1.9%
No.6	横ばい	下降	92	5.3%
No.7	下降	上昇	194	11.2%
No.8	下降	横ばい	204	11.8%
No.9	下降	下降	824	47.7%

増加(↗)パターン: 30%
 減少(↘)パターン: 70%

新商品導入後3週間のデータから回帰分析により予測しようとするとき大きく外れる可能性

Appendix

▶ ブームピーク後の解

$$y_1(t) = (1 - k)S \cdot e^{-b_1(t-T)},$$

$$y_2(t) = \frac{b_1(1 - k)S}{b_1 - b_2 - b_3} \left\{ -e^{-b_1(t-T)} + e^{-(b_2+b_3)(t-T)} \right\},$$

$$y_3(t) = \frac{b_2(1 - k)S}{b_2 + b_3} + \frac{b_2(1 - k)S}{(b_2 + b_3)(b_1 - b_2 - b_3)} \left\{ (b_2 + b_3)e^{-b_1(t-T)} - b_1e^{-(b_2+b_3)(t-T)} \right\},$$

$$y_4(t) = \frac{(b_2k + b_3)S}{b_2 + b_3} + \frac{b_3(1 - k)S}{(b_2 + b_3)(b_1 - b_2 - b_3)} \left\{ (b_2 + b_3)e^{-b_1(t-T)} - b_1e^{-(b_2+b_3)(t-T)} \right\}$$

$$Y(t) = \frac{\{(b_2t + b_3T)k + b_3(t - T)\} bS}{b_2 + b_3} + \frac{(1 - k)bS}{b_1 - b_2 - b_3} \left\{ \frac{b_3 - b_1}{b_1} (1 - e^{-b_1(t-T)}) + \frac{b_1 b_2}{(b_2 + b_3)^2} (1 - e^{-(b_2+b_3)(t-T)}) \right\}.$$

Appendix

▶ ブームのピーク

▶ 「製品の消費顧客数 $y_2(t)+y_4(t)$ が極大値を取るとき」

ブームピーク時刻 T^*

$$T^* = T + \frac{\ln\left(\frac{b_1 - b_3}{b_2}\right)}{b_1 - b_2 - b_3} \quad \dots(13)$$

$b_1 - b_3 < 0$ であるとブームのピークは存在しない

Appendix パラメータの弾力性

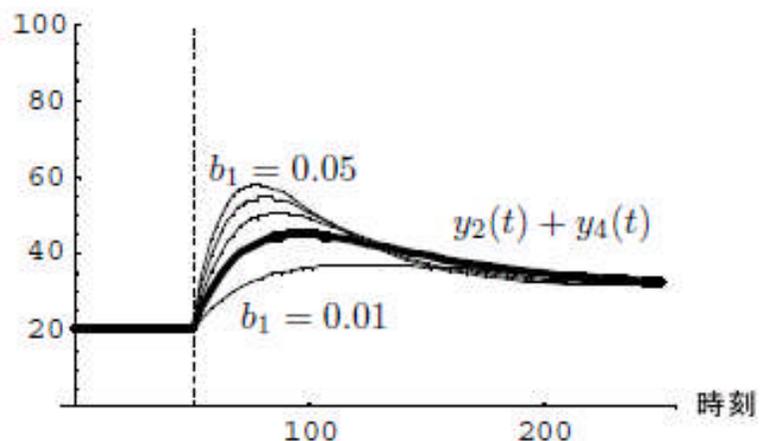


図10(a). b_1 :0.01~0.05 まで0.01刻み

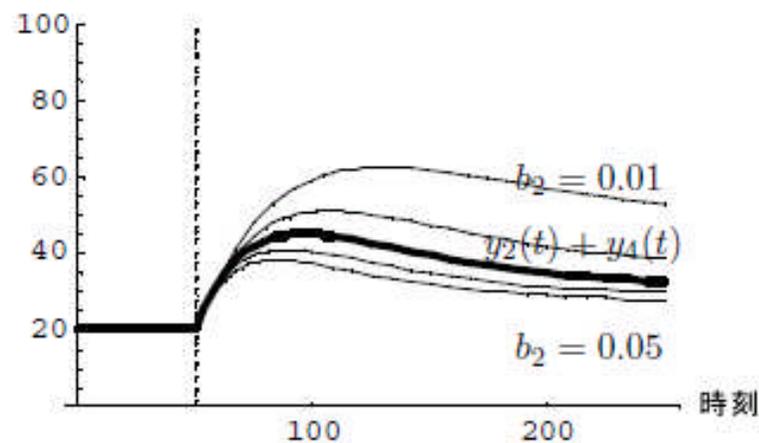


図10(b). b_2 :0.01~0.05 まで0.01刻み

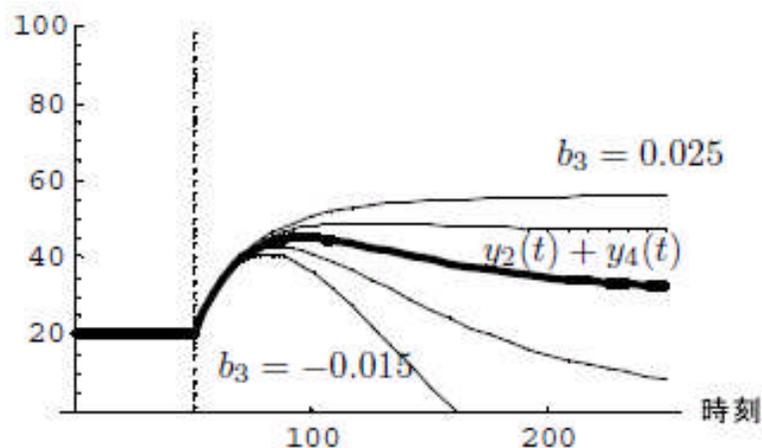


図10(c). b_3 :-0.015~0.025 まで0.01刻み

図10. ブームモデル消費顧客数
各パラメータに関する感度

Appendix 期間と売上パターンとの関係

▶ ワイブル分布はパラメータの値から分布の形が変わる

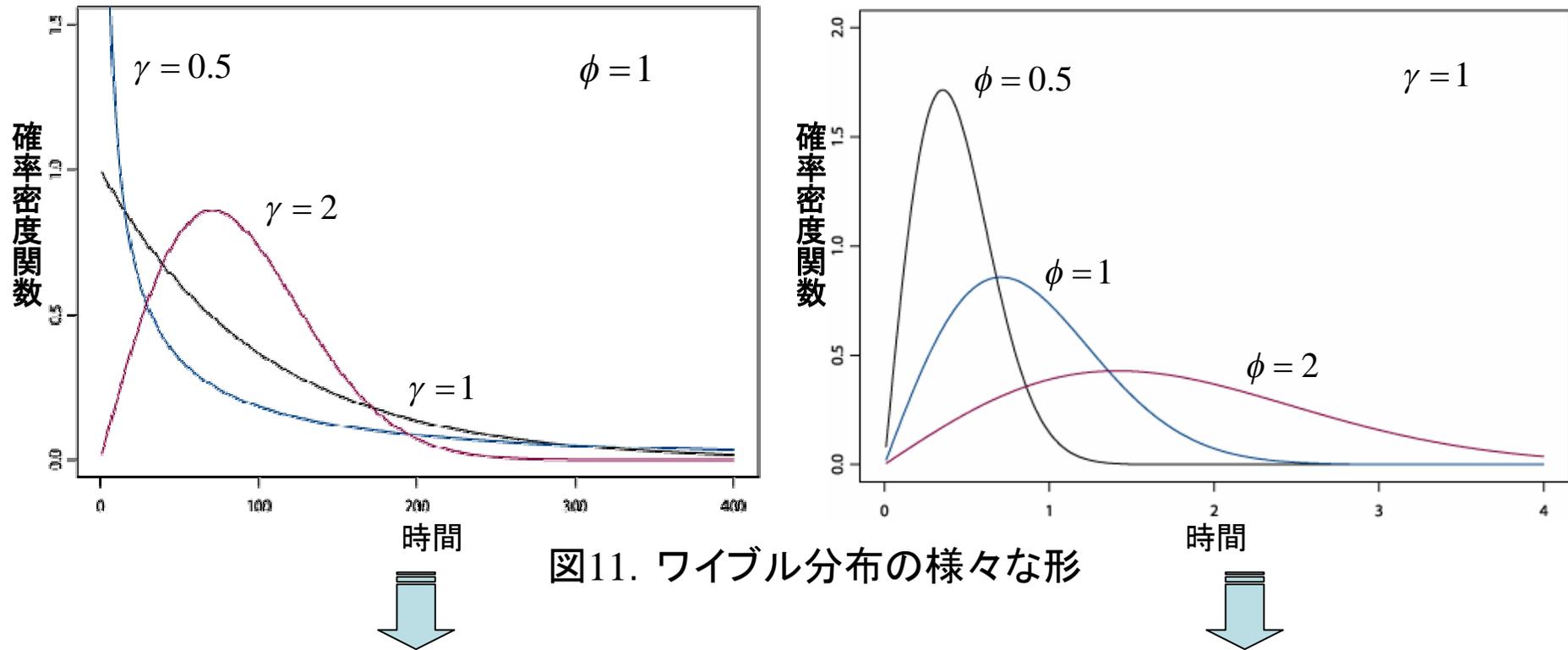


図11. ワイブル分布の様々な形

γ : 1より大きい→販売日から少し経過して売れる
 1より小さい→販売日より近くで売れる

ϕ : 大きい→長い期間売れる
 小さい→短期間で売れる

Appendix ベイズの利点[8]

1. どのような状況下でもベイズの定理は例外無く適用される
2. モデルが与えられれば, ベイズ推測ではデータを含む情報を自動的に活用できる
3. データ以外の知識や情報を活用できる
4. 漠然とした事前分布を利用可
5. 未知量について直接確率計算が行える
6. データや情報の蓄積を自然に容易に活用できる
(簡単に更新ができる)
7. 母数について制約があっても問題が生じにくい