顧客生涯価値に関する実証研究
– 株式会社ジィ・シィ企画 提出用報告書 –

北海道大学大学院経済学研究院*

June 30, 2019

*本報告書についての問い合わせは，北海道大学大学院経済学研究院 地域経済経営ネットワーク研究センター（REBN, 連絡先 sacade@econ.hokudai.ac.jp）までお願いいたします．

はじめに

はじめに
概要と結論
パレート則・RFM指標
データと結果の概要
顧客行動のモデル化と推定
結果の考察
結論と課題

概要と結論

- 大規模データの蓄積に基づくマーケティング： 顧客管理への活用
  - パレート則：売上の８割を２割の「優良」顧客が占めている
  - 帰結：上位優良顧客への集中
- 過去の売り上げの大きさ＝優良？
  - 購入金額のみで見る（Figure 1）
  - 購買頻度や直近購買を考慮（Figure 2，左・中パネル）
- 店舗あるいは商店街と顧客の関係に基づく価値＝顧客生涯価値（Customer Lifetime Value; CLV）
  - 顧客生涯価値：購入金額の割引現在価値（ただし顧客が消えていくことも考慮の必要あり）
  - 要約統計量： 直近 (Recency) ・頻度 (Frequency) ・金額 (Monetary)
  - 顧客行動のモデル化により，RFM から CLV を推定

概要と結論（つづき）

- 結論：来訪の途絶えた顧客の考慮 → 集中度はマイルドに（Figure 2，右パネル）
  - パレート則よりも集中は弱まり，上位 10%で約 40%，上位 20%で 60%の総価値
    - 金額ベースのみでは，上位 10%で約 60%，上位 20%で 80%の総購入額
  - 上位顧客のみにマーケティング資源を集中させると，大きな CLV をもたらす顧客を見過ごす危険
    - 直近の $5 \sim 10$ 日以内に商店街へ来訪，来訪頻度の高かった顧客（Table 7）
    - 「購買額／日」や「購買額／件」は必ずしも劇的に高くない（Table 6 と Table 7）
- 顧客が長期的に足繁く商店街へ通う仕組みこそ，商店街へもたらされる顧客生涯価値の総額を増加させ，長期的な利益につながる可能性

ROC 曲線：購入額の集中度

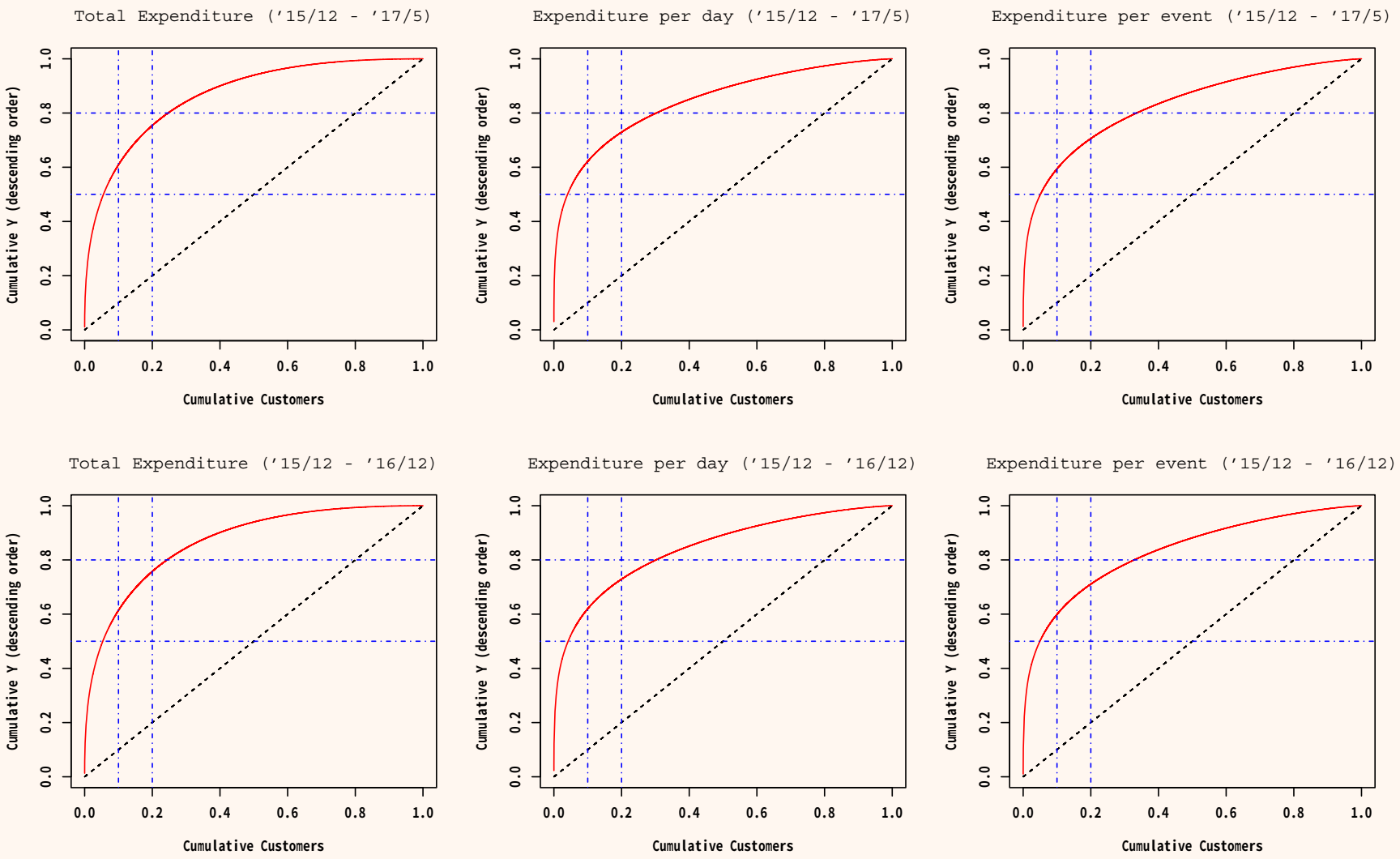

Figure 1: **上位購買顧客への集中：総購入額，購入額／日，購入額／件**

ROC 曲線：購入額の集中度

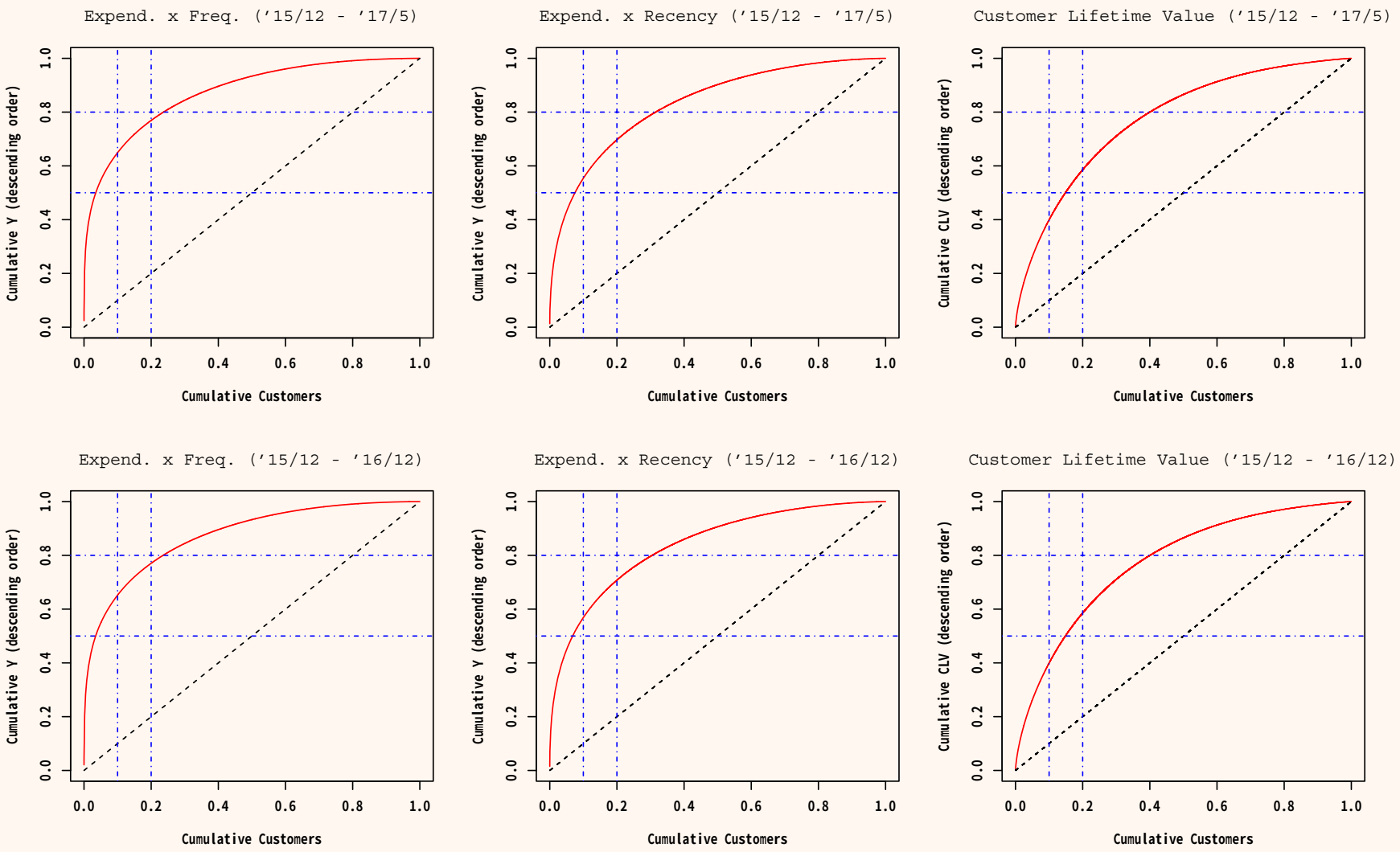

Figure 2: **上位購買顧客への集中度：購入頻度・直近購入の考慮と** CLV

パレート則・RFM 指標

- 上位 10%の顧客への集中が顕著（Figure 1）
  - 典型的には，上位 10%の顧客が，全体の 60%を占めている
  - 上位 20%の顧客が，全体の 80% となっているのは総購入額
  - 他の購入額指標では，上位 20%の顧客が，全体の約 70%
- 離脱者の可能性
  - 過去において高額購入を行っていても，すでに離脱している可能性があれば，今後は「良い」顧客とは言えない
  - 顧客が離脱するまでの期間を考慮し，今後支出される金額の（割引現在）価値によって,「良さ」を評価できないか

パレート則・RFM 指標

- 購入頻度と直近購入の考慮（Figure 2）
  - 今も「離脱していない」顧客
    - 基準１：週に一回以上購入を行っている
    - 直近の購入が 15 週（約三か月）以内である
  - 上記の基準を満たす顧客を選別し，さらに上位顧客への集中度を確認
    1. 上位 10%の顧客への集中は頑健（Figure 2の左から 2 列のパネル）
       - 上位 10%の顧客が，全体の 60%
       - 上位 20%の顧客が，全体の 70〜80%

データと結果の概要

はじめに
データと結果の概要
データ：データセットの概要
データ：標本特性
記述統計：全体の平均・分散など
店舗別購買タイミングのパターン
顧客行動のモデル化と推定
結果の考察
結論と課題

データ：データセットの概要

- 記録期間：2014 年 12 月 1 日から 2017 年 5 月 31 日
  - 2,049,122 レコード
  - 68,001 人分の顧客購買（およびポイント交換）履歴
  - 会員番号が分からないレコードは除外.
- 標本に含める顧客
  - 2015 年 12 月末日までに最初の取引のある顧客：40,736 人
    - 顧客行動のモデル化：一定期間の履歴が必要
    - 購買活動を行っているレコードのみに限定（ポイント交換のみを行っているレコードは除外）
- モデルの推定・モデルの検証・最終的な CLV 推定
  - モデルの推定：2014 年 12 月末日から 2016 年 12 月末日のレコード
  - 予測によるモデルの検証：2017 年１月１日から 2017 年５月末日のレコード
  - 最終的な推定：2014 年 12 月末日から 2017 年５月末日のレコード

Table 1: **記述統計：店舗別レコード数（2014 年 12 月から 2017 年 5 月）**

店舗グループ	レコード数		店舗グループ	レコード数
食品	641,020		リラクゼーション	46,251
健康食品	322,815		眼鏡	39,276
服飾品	221,227		レストラン	39,210
靴	44,930		薬局	37,322
カフェ	218,560		スポーツ	25,950
雑貨	169,065		きもの	11,247
電器	97,739		宝石	9,072
その他	68,875		携帯電話	2,677
サービスカウンター	53,886			
			合計	2,049,122

データ：標本特性

- 購買額（Monetary）
  - 顧客ごとの総購買額
  - 顧客ごとの購買額／買物日
  - 顧客ごとの購買額／購入１件
- 購入頻度（Frequency）
  - 購買日数：何日商店街で買い物したか
  - 購買日数／週：最初の取引記録から観測終了までに，週当たり何日買い物へ行ったか
- 直近購入（Recency）
  - 観測終了時点から一番直近までの購入時点
- 店舗種類別の利用割合
  - 購買履歴において，どの店舗で購買したレコードがどの程度の割合を示すか
  - 顧客の様々な特性の現れ

記述統計：全体の平均・分散など

Table 2: 記述統計：2014 年 12 月から 2017 年 5 月

	平均	標準偏差	最小値	10%点	50%点	90%点	最大値
年齢	55.597	17.210	2	33	54	79	114
性別（女性）	0.781	0.413	0	0	1	1	1
購買日数	35.286	49.187	1	2	18	88	737
購買日数/週	0.275	0.373	0.008	0.018	0.146	0.677	5.506
直近購買（週）	29.062	38.904	0	0.429	7.857	99.714	134.00
購買総額（万円）	19.790	86.367	0.0001	0.590	6.399	38.978	8897.567
購買/日（千円）	8.519	79.566	0.001	1.173	3.145	12.393	10570.671
購買/件（千円）	6.275	36.439	0.001	1.001	2.503	9.489	3171.201
食品店舗（%）	0.262	0.256	0.000	0.000	0.200	0.636	1.00
健康店舗（%）	0.121	0.201	0.000	0.000	0.013	0.400	1.00
カフェ（%）	0.099	0.155	0.000	0.000	0.033	0.286	1.00
衣料店舗（%）	0.182	0.226	0.000	0.000	0.104	0.500	1.00

データ：データセットの概要

- 記述統計（Table 2）より
  - 顧客の構成員は，女性で，年齢は 55.6 歳以上に偏っている．
  - 平均的には月に一度（４週に一度）は商店街で購買活動を行っている．
  - 週数で表現した直近購買は比較的大きく，メディアンで 7.9 週（50 日）程度．
  - いったん商店街来ると，メディアンで 3 千円以上を支出（一回の買い物では約 2.5 千円）．
  - 主に食料品を扱う店舗，衣料品・履物を扱う店舗での買い物件数が多い．

店舗別購買タイミングのパターン

Figure 3: **購買タイミング（無作為 30 名）**

店舗別購買タイミングのパターン

- 全店舗：標本期間内に閉店した店舗を含め 129 店舗の取引記録
  - 全体の購買タイミング：多くのパターンや嗜好をもつ顧客の混在
- 類似性の高く，レコード数も多い４つのグループを抽出
  - 食品を主に取りつかう店舗
    - 顧客にとってのアクセスの良さの指標
  - 健康食品を主に取りつかう店舗
    - 強い好みを反映，通い続けるか，試した後通わないかが極端に出る
  - カフェを営んでいる店舗
    - 休息や娯楽のため，思いついた時あるいは定期的に利用する傾向
  - 衣類・履物類を主に販売している店舗
    - 必要な時に利用する可能性が高そうである

店舗別購買タイミングの規則性パラメータ k

- 来訪パターン＝訪問間隔時間：ガンマ分布（$\Gamma(k, k\theta)$）のパラメータ k
  - $k < 1$：（時折）まとまって事態が生じるパターン (clumpy)
  - $k = 1$：ランダムに事態が生じるパターン (random) ＝指数分布
  - $k > 1$：規則正しく事態が生じるパターン (clumpieness)
- このパラメータ k を各顧客について推定し，箱ひげ図としてプロット（Figure 4）
  - 縦の黒線は，顧客ごとに推定した k のメディアン（50%点）
  - 大部分の推定値は１近辺に位置している（タイミングはランダム）
    - ０に近いところ（ある時点でまとめて来るものの，しばらく来ない）
    - ２を越える（規則的なタイミングで来る）大きな推定値も見られる
  - 異質性も無視できない
    - 全顧客が１に近く均質なら，以下のモデルの設定（指数分布）は妥当
    - k が顧客間で異質であるなら，モデルの拡張の必要性

店舗別購買タイミングの規則性パラメータ k

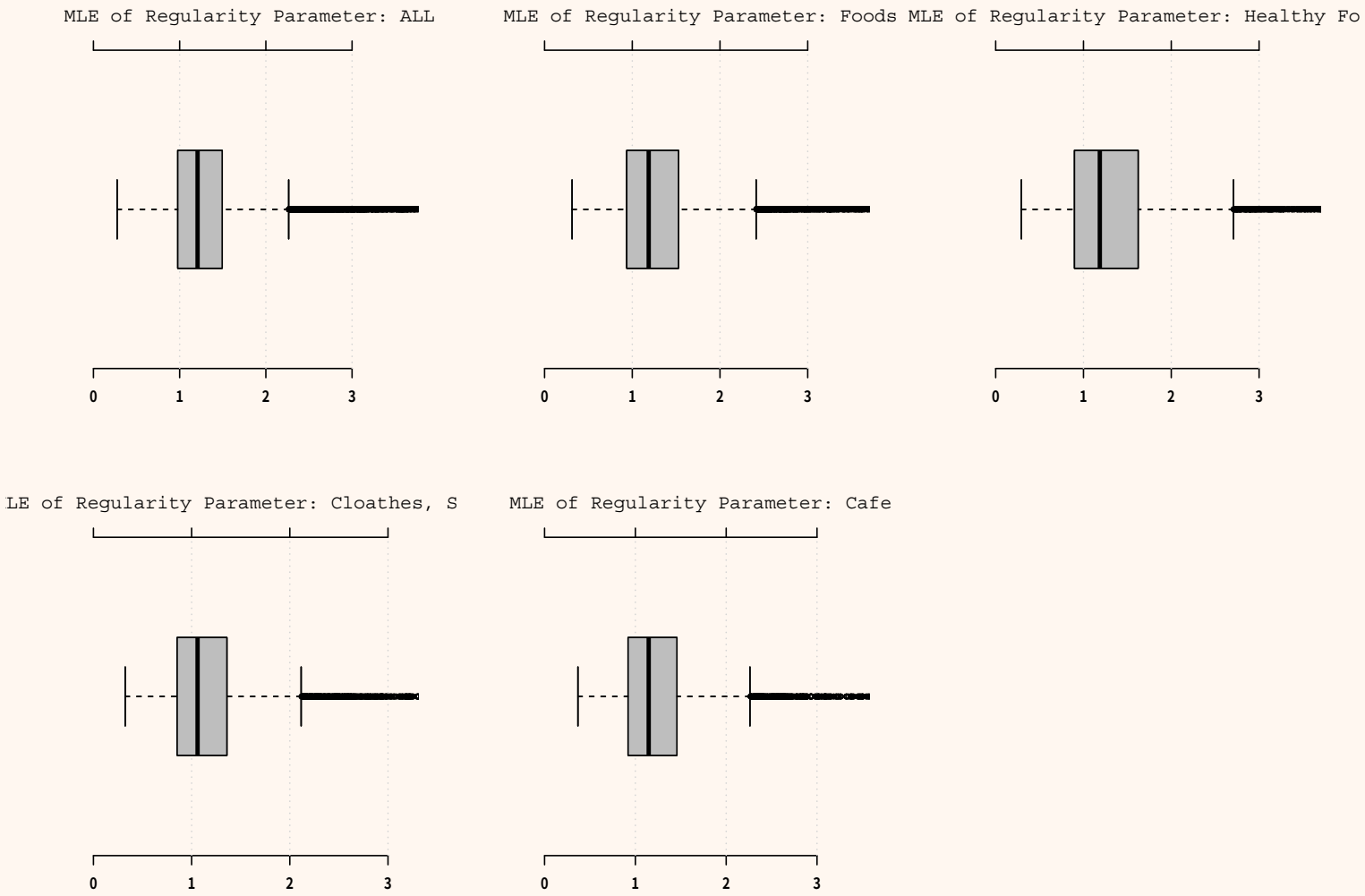

Figure 4: 各顧客の購買タイミングに関する規則性パラメータ推定値

店舗別購買タイミングの規則性パラメータ k

- どこかの店舗（全店舗）での購買：規則性パラメータの平均＝ 1.208
  - 顧客ごとの来訪パターンが，利用店舗や嗜好によって混在．
  - 何らかの形で定期的に利用する人も多いため，k ＞１の傾向はあるが，多くの顧客が１近辺に分布．
    - 次節のモデルは，顧客間で均質に $k = 1$ であることが前提．
- 食品店舗での購買：規則性パラメータの平均＝ 1.187
- 健康食品店舗での購買：規則性パラメータの平均＝ 1.186
  - 定期的に通う顧客（k が１よりかなり大きい）が比較的多い
  - k がは２を越えるほど強い定期性を示す顧客も多い
- カフェ型店舗での購買：規則性パラメータの平均＝ 1.147
- 衣類・履物店舗での購買：規則性パラメータの平均＝ 1.061
  - 定期的に通う顧客の一方で，ランダムなパターン（$k \approx 1$）で来訪

はじめに
データと結果の概要
顧客行動のモデル化と推定
モデル
モデルの推定
結果の考察
結論と課題

顧客行動のモデル化と推定

モデル

1. **顧客同士は独立に行動する（**$t = 1, 2, \ldots, T_i,\ i = 1, 2, \ldots, n.$ **）**
2. **顧客** i **の一日当たりの購買額は対数正規分布に従う（平均は** $\exp\{\eta_i + 0.5\omega^2\}$, **分散は）．**

$$\log s_{i,t} \sim N(\log \eta_i, \omega^2), \quad t = 1, 2, \ldots, T_i, \quad i = 1, 2, \ldots, n.$$

3. **顧客** i **の離脱までの時間は，独立に指数分布に従う（パラメータは** μ_i**）**

$$\Pr[\mathcal{T} \leq \tau_i] = 1 - e^{-\mu_i \tau_i}, \quad \tau_i \geq 0.$$

モデル

4. 顧客 i の購買間隔は，独立に指数分布に従う（パラメータは λ_i）

$$\Pr[dT_j \leq \Delta t_{j,i}] = 1 - e^{-\lambda_i \Delta t_i}, \quad \tau_i \geq 0.$$

ただし，

$$0 = t_{0,i} < t_{1,i} < t_{2,i} < \cdots < t_{x_i-1,i} < t_{x_i,i} = T_i.$$

次の瞬間の購入強度＝ハザード関数：今の特定化の下ではどの時点でも一定（λ_i）．

- 仮定４は，時点 0 において最初の購買がなされた後，期間 $(0, T_i]$ における購買頻度が，購買強度 λ_i のポアソン分布に従うことを含意

$$\Pr[X = x_i] = \begin{cases} \frac{(\lambda_i T_i)^{x_i} \exp\{-\lambda_i T_i\}}{x_i} & \text{if } \tau_i > T_i \\ \frac{(\lambda_i T_i)^{x_i} \exp\{-\lambda_i T_i\}}{x_i} & \text{if } \tau_i \leq T_i \end{cases}, \quad x_i = 0, 1, \ldots$$

- 購買タイミングに強い異質性があるとき，仮定４は妥当でない．

モデル

5. λ_i, μ_i はそれぞれ購買強度，離脱率を表し，その水準は個体ごとに異なっている．またいくつかの外生変数に依存する形に特定化することもできる

$$\begin{aligned}\log \lambda_i &= \beta_{\lambda,0} + \beta_{\lambda,1} X_{1,i} + \cdots + \beta_{\lambda,K} X_{K,i} + U_{\lambda,i} \\ \log \mu_i &= \beta_{\mu,0} + \beta_{\mu,1} X_{1,i} + \cdots + \beta_{\mu,K} X_{K,i} + U_{\mu,i}\end{aligned}$$

- 購入強度および離脱率と関連付ける外生変数としては以下のものを用いた
  - 年齢（ごく少数欠損値が存在したが，平均値で代用した）
  - 性別（女性が１，男性は０）
  - 購買イベントに占める食品購入割合 （顧客にとって，定期的に利用する上で，商店街の利便性についての指標）
  - 購買イベントに占める健康食品割合 （顧客にとって，自らの嗜好を反映し，定期的に利用する上での忠実さの指標）
  - 購買イベントに占めるカフェ割合 （顧客にとって，自らの嗜好を反映し，ランダムに利用する上での忠実さの指標）
  - 購買イベントに占める衣料品購入割合（顧客にとって，ランダムに利用する上で，商店街の利便性についての指標）

モデルの推定

- ベイズ的統計手法の利用
  - 顧客ごとの異質性（λ や μ が顧客ごとに異質であること）があり，個々の λ や μ にも興味がある
  - その他に λ や μ を用いた統計量の分布にも興味がある
  - ただし，事後分布を陽表的に求められないので，マルコフ連鎖モンテカルロ法（メトロポリス・ヘイスティング法）を用いて，各パラメターに関する事後分布を発生させ，推定結果の評価を行う
- 推定手順：阿部 (2011) によるマルコフ連鎖モンテカルロ法
  - ただし，顧客 i 購買額の平均 η_i と共通の分散 ω^2 については，購入強度および離脱率の式とは別個に回帰分析によって推定した．

モデルの推定

- 推定に用いた標本
  - 2015 年 12 月 31 日以前に最初の購入が記録されている顧客
  - 観測期間は 2016 年 12 月 31 日まで
    - 残り 5 か月分は推定結果の予測力の評価に用いるため推定では除外
  - 観測対象となったのは 40,736 名の顧客
- モデルの予測力
  - 外挿期間（2017 年 1 月 1 日から 5 月 31 日まで）における購買行動をそれ以前のデータを用いて予測する
  - 外挿期間の「購買回数」は観測可能
  - 「商店街から完全に離脱したかどうか」は観測不可能
    - 「外挿期間（５か月分）に購買有」＝「離脱してない」は観測可能
  - 「終了後５か月間生存している確率」あるいは「終了後５か月以内に購買を行う確率」を推定

モデルの推定

Table 3: 外挿期間における予測力の評価

	頻度との相関	活動との相関	生存との相関
説明変数あり	0.875	0.813	0.846
説明変数なし	0.874	0.833	0.846

- モデルの予測力
  - 外挿期間（2017 年 1 月 1 日から 5 月 31 日まで）における購買行動をそれ以前のデータを用いて予測
    - 「外挿期間（５か月分）の購買頻度」と「５か月分予測購買頻度」の相関係数
    - 「外挿期間（５か月分）の購買の有無」と「終了後５か月以内に購買を行う確率」の相関係数
    - 「外挿期間（５か月分）の購買の有無」と「終了後５か月間生存している確率」の相関係数

推定結果（1）

Table 4: **推定結果：**2014 **年** 12 **月から** 2016 **年** 12 **月**

係数	事後分布分位点 2.5%	50%	97.5%		事後分布分位点 2.5%	50%	97.5%	
$\beta_{\lambda,0}$	-2.54	-2.49	-2.45	*	-1.86	-1.84	-1.83	*
$\beta_{\lambda,age}$	-21.54	-0.11	19.87		–	–	–	
$\beta_{\lambda,gen}$	0.13	0.16	0.18	*	–	–	–	
$\beta_{\lambda,HL}$	1.34	1.42	1.48	*	–	–	–	
$\beta_{\lambda,FD}$	1.23	1.29	1.34	*	–	–	–	
$\beta_{\lambda,CS}$	-0.86	-0.80	-0.72	*	–	–	–	
$\beta_{\lambda,CF}$	1.47	1.57	1.65	*	–	–	–	
$\beta_{\mu,0}$	-5.17	-5.08	-4.98	*	-6.20	-6.13	-6.09	*
$\beta_{\mu,age}$	-22.35	0.09	23.11		–	–	–	
$\beta_{\mu,gen}$	-0.36	-0.30	-0.23	*	–	–	–	
$\beta_{\mu,HL}$	-0.86	-0.73	-0.59	*	–	–	–	
$\beta_{\mu,FD}$	-1.48	-1.34	-1.22	*	–	–	–	
$\beta_{\mu,CS}$	-2.28	-1.85	-1.55	*	–	–	–	
$\beta_{\mu,CF}$	-1.75	-1.54	-1.34	*	–	–	–	
σ^2_λ	1.12	1.14	1.16	*	1.37	1.39	1.42	*
$\sigma_{\lambda\mu}$	-0.74	-0.71	-0.67	*	-0.91	-0.87	-0.82	*
σ^2_μ	1.09	1.42	1.68	*	1.26	1.42	1.71	*

推定結果（2）

Table 5: **推定結果：**2014 **年** 12 **月から** 2017 **年** 5 **月**

	事後分布分位点				事後分布分位点			
係数	2.5%	50%	97.5%		2.5%	50%	97.5%	
$\beta_{\lambda,0}$	-2.60	-2.56	-2.52	*	-1.88	-1.87	-1.85	*
$\beta_{\lambda,age}$	-20.15	0.00	24.32	–	–	–	–	
$\beta_{\lambda,gen}$	0.13	0.15	0.18	*	–	–	–	
$\beta_{\lambda,HL}$	1.40	1.46	1.53	*	–	–	–	
$\beta_{\lambda,FD}$	1.29	1.34	1.39	*	–	–	–	
$\beta_{\lambda,CS}$	-0.75	-0.68	-0.60	*	–	–	–	
$\beta_{\lambda,CF}$	1.55	1.65	1.74	*	–	–	–	
$\beta_{\mu,0}$	-5.40	-5.32	-5.24	*	-6.29	-6.18	-6.10	*
$\beta_{\mu,age}$	-20.58	-0.29	17.36		–	–	–	
$\beta_{\mu,gen}$	-0.35	-0.30	-0.25	*	–	–	–	
$\beta_{\mu,HL}$	-0.71	-0.57	-0.42	*	–	–	–	
$\beta_{\mu,FD}$	-1.17	-1.05	-0.93	*	–	–	–	
$\beta_{\mu,CS}$	-1.05	-0.86	-0.65	*	–	–	–	
$\beta_{\mu,CF}$	-1.44	-1.28	-1.10	*	–	–	–	
σ^2_λ	1.16	1.18	1.20	*	1.41	1.44	1.46	*
$\sigma_{\lambda\mu}$	-0.67	-0.64	-0.60	*	-0.82	-0.78	-0.73	*
σ^2_μ	0.88	0.95	1.11	*	0.87	1.15	1.57	*

サンプリング結果（1）

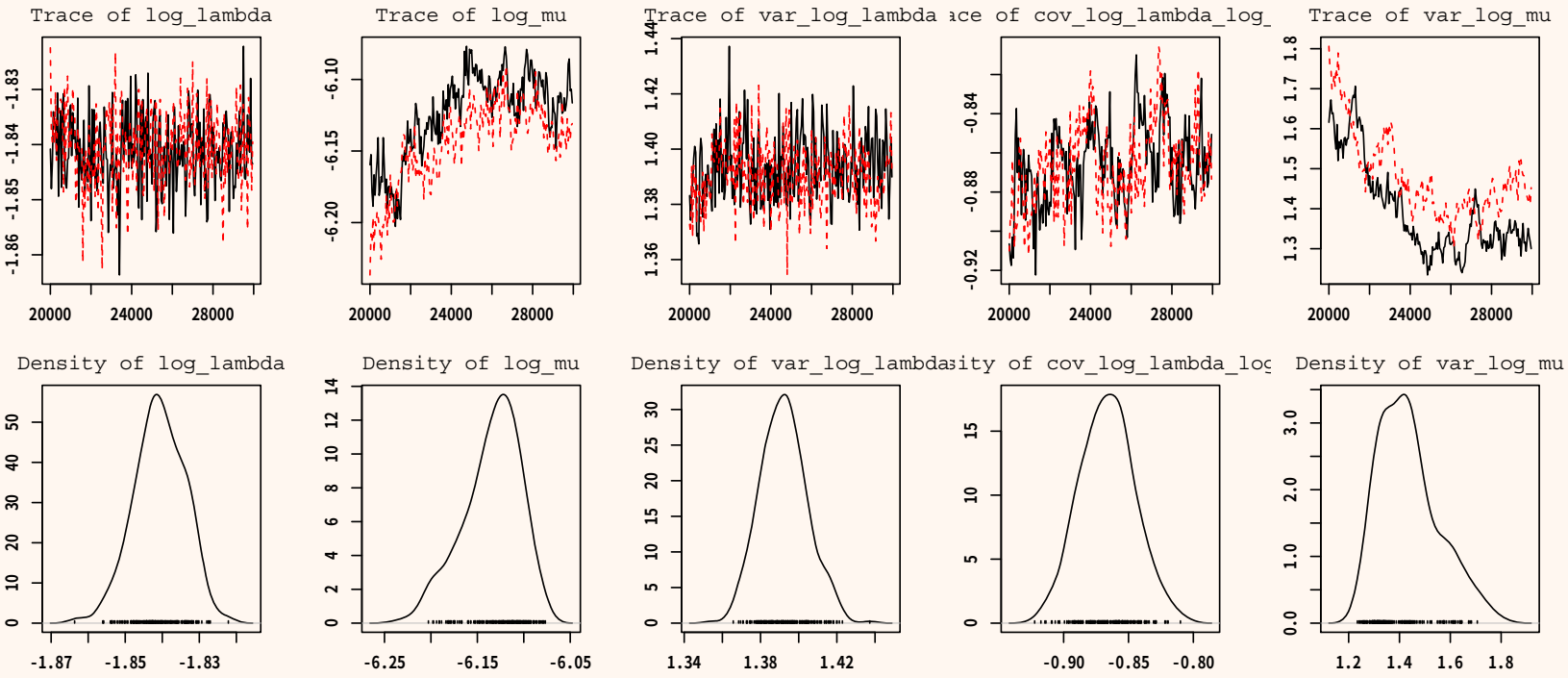

Figure 5: MCMC のサンプリング結果（説明変数なし，2014 年 12 月から 2016 年 12 月まで）

サンプリング結果（２－１）

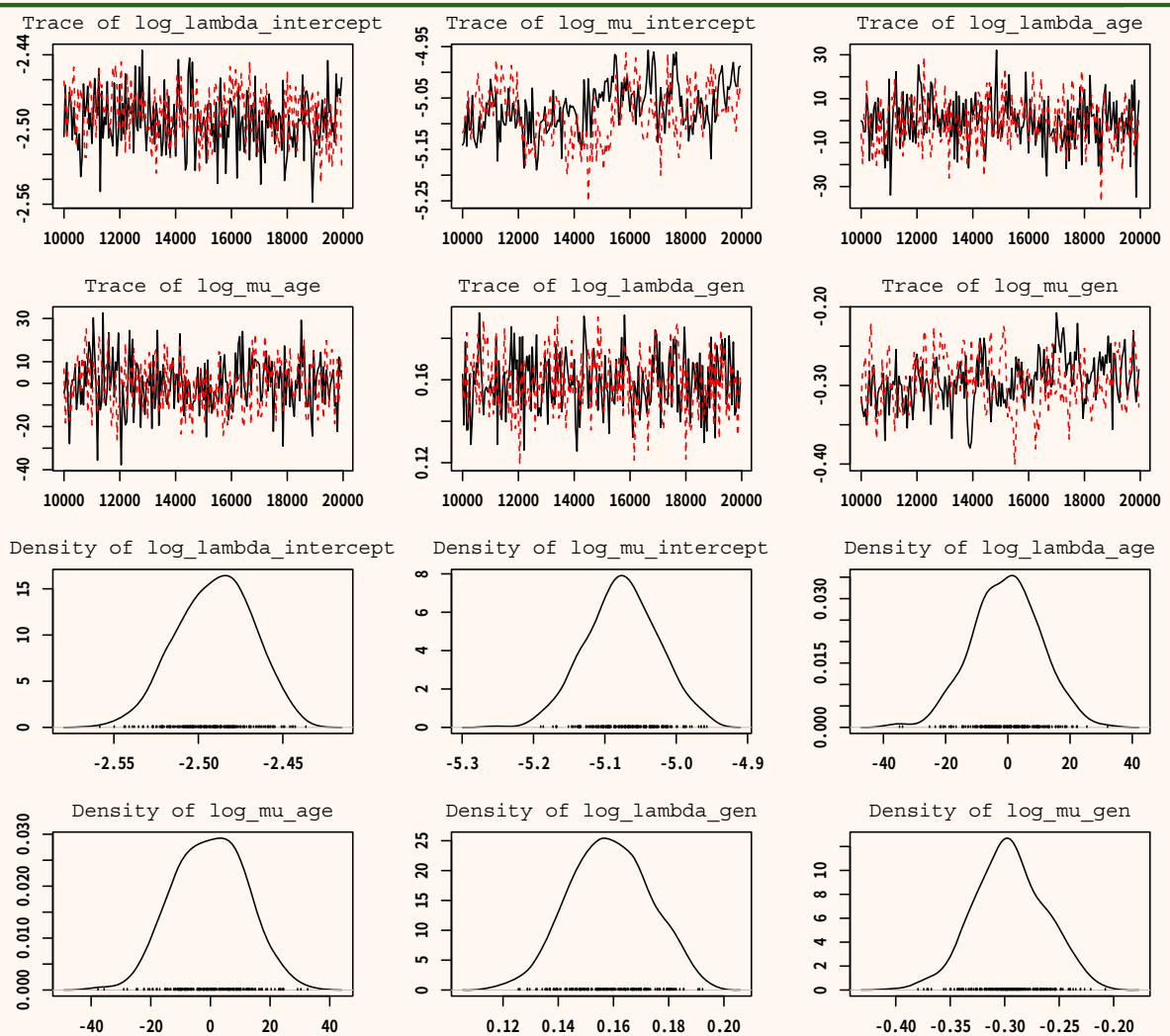

Figure 6: MCMC のサンプリング結果（説明変数あり，2014 年 12 月から 2016 年 12 月まで）

サンプリング結果（2－2）

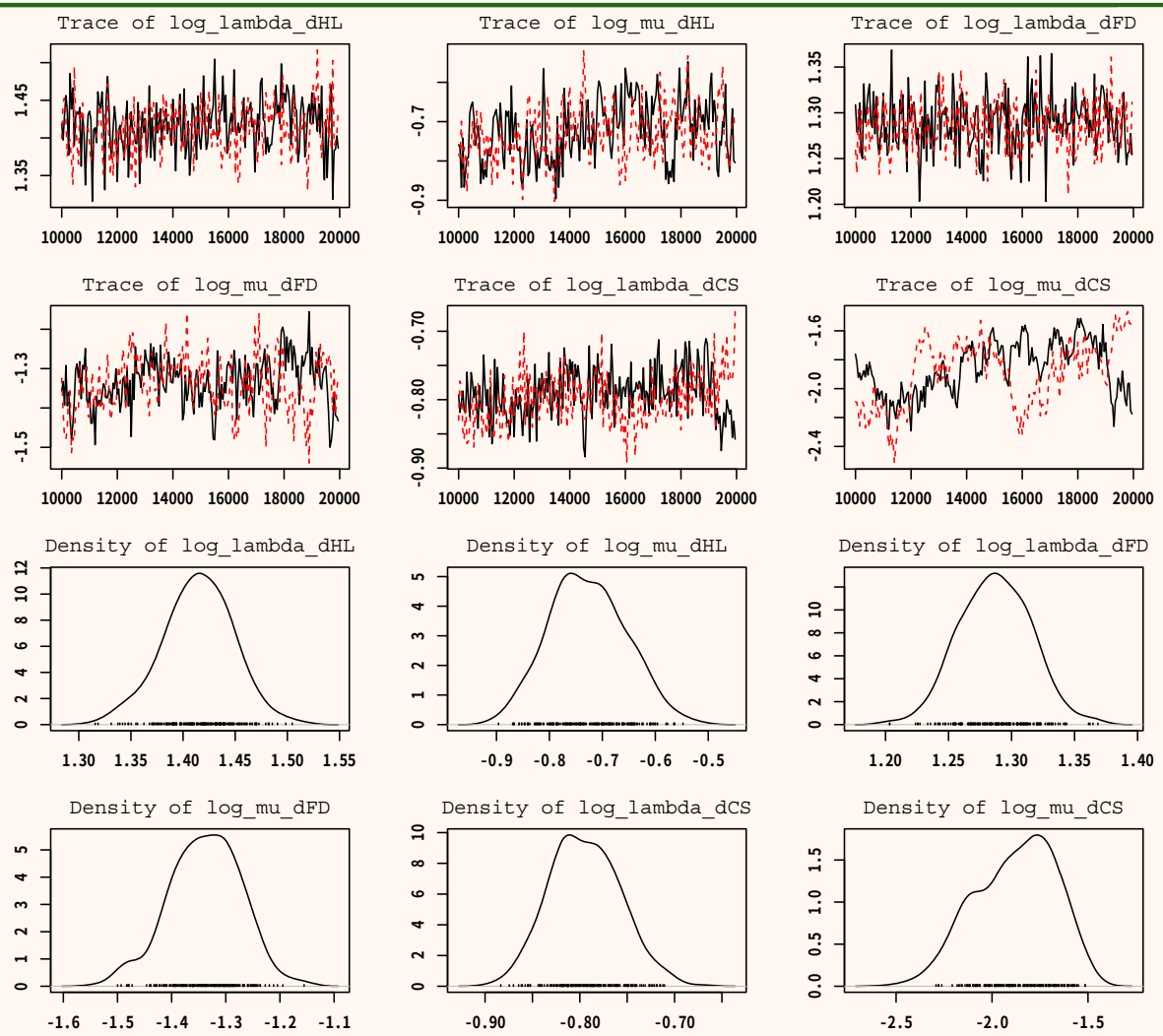

Figure 7: MCMC のサンプリング結果（説明変数あり，2014 年 12 月から 2016 年 12 月まで）

サンプリング結果（2－3）

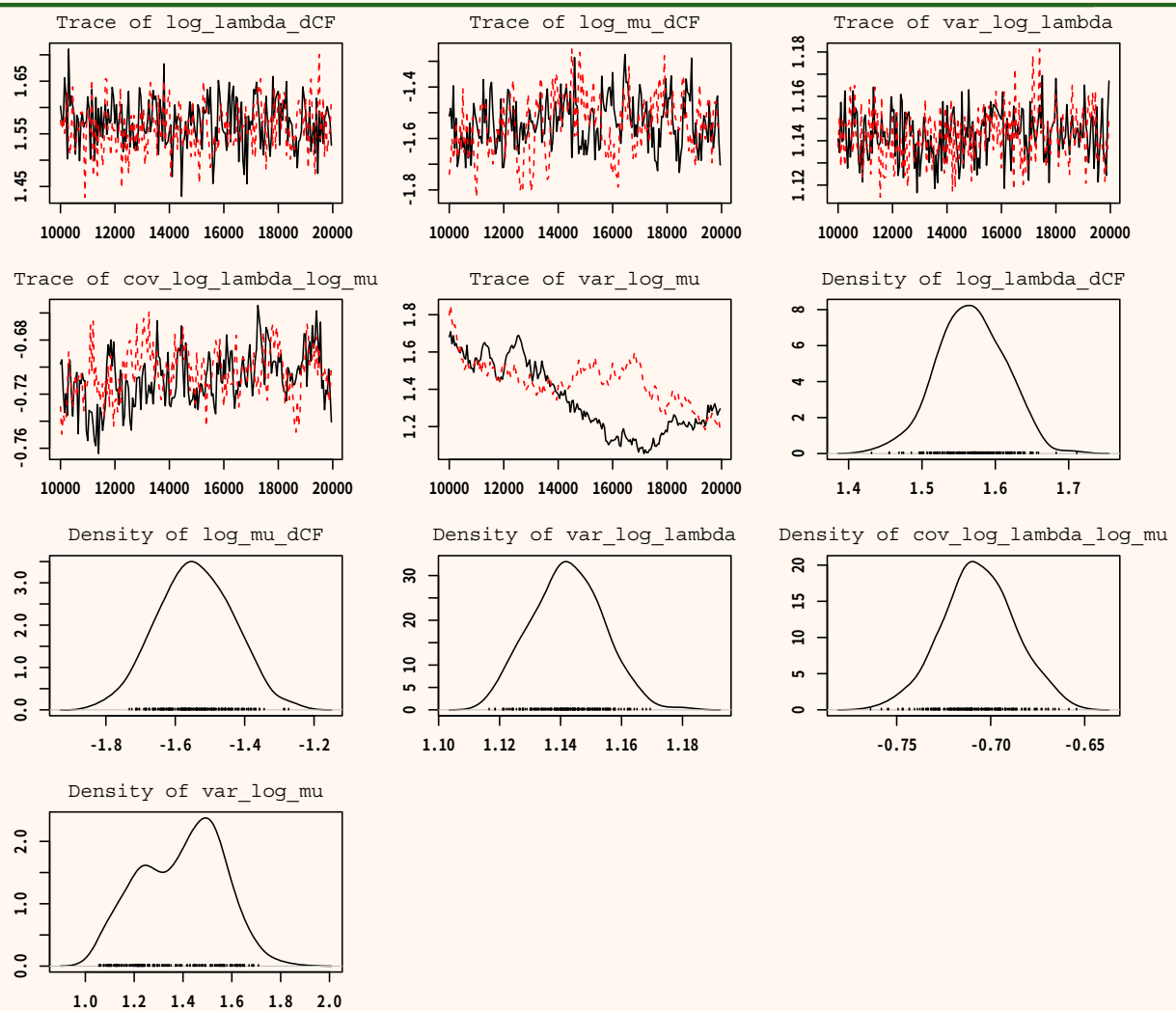

Figure 8: MCMC のサンプリング結果（説明変数あり，2014 年 12 月から 2016 年 12 月まで）

結果の考察

はじめに
データと結果の概要
顧客行動のモデル化と推定
結果の考察
推定結果に基づいた顧客生涯価値の特性
考察
結論と課題

推定結果に基づいた顧客生涯価値の特性

- 顧客生涯価値：生存している顧客の生み出す購買額流列の割引現在価値
  - 時点 t の顧客 i の顧客価値：
    - $V_i(t)$ ＝時点 t の購入強度 $_i$ $\times$ 期待購入額 $_{i,t}$ ＝ $\lambda_i \times \eta_i e^{\omega^2/2}$（時点不変）.
  - 時点 t の顧客 i の生存確率：$R_i(t) = e^{-\mu_i t}$.
  - 時点 t の顧客 i の割引率：$D_i(t) = e^{-\delta t}$（年率 10 %として $\delta = \log(1+0.1)$）.

$$
\begin{aligned}
\mathsf{CLV}_i &= \int_0^\infty V_i(t) R_i(t) D_i(t) dt \\
&= \int_0^\infty \left\{ \lambda_i \eta_i e^{\omega^2/2} \right\} e^{-\mu_i t} e^{-\delta t} dt = \frac{\lambda_i \eta_i}{\mu_i + \delta} \exp\left\{ \frac{\omega^2}{2} \right\}.
\end{aligned}
$$

CLV と総購買額

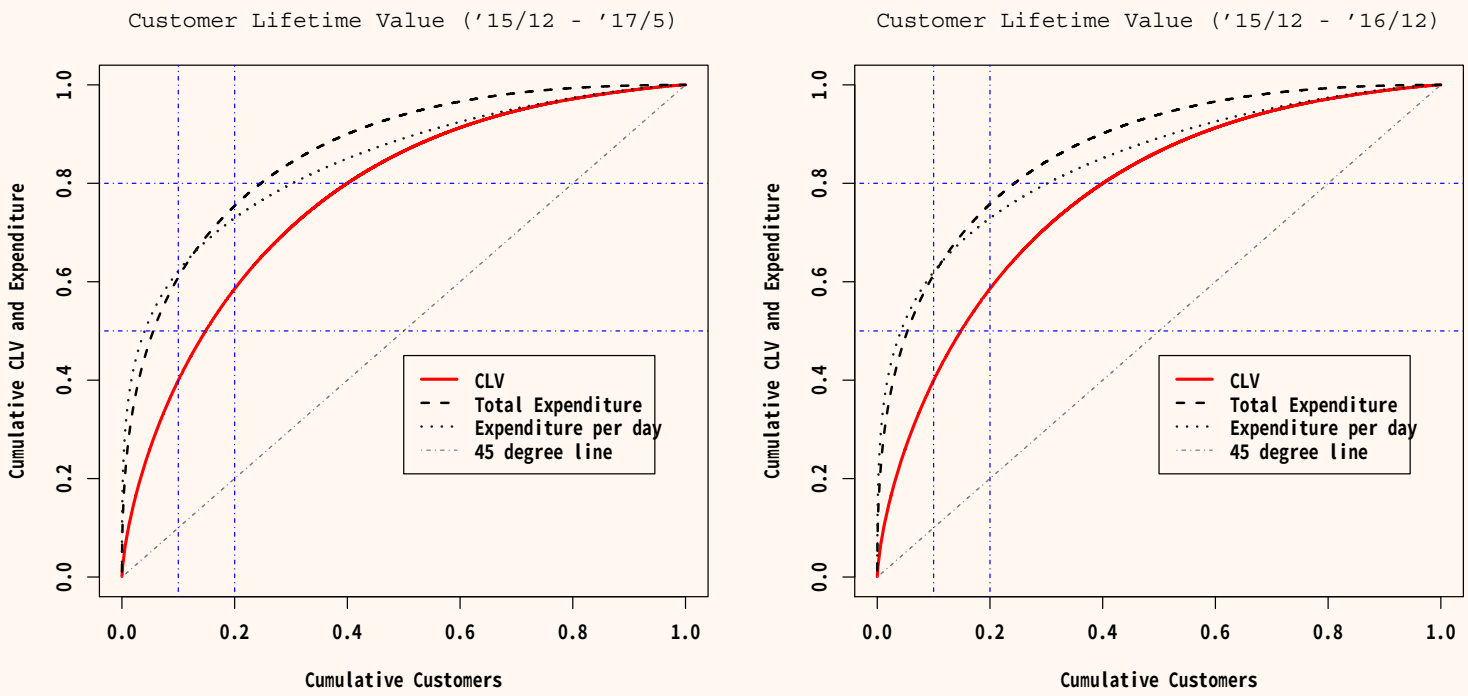

Figure 9: ROC 曲線：顧客生涯価値と総購買額

- **ジニ係数による比較：**
  - **顧客生涯価値の** ROC **曲線：** 0.550 ／ 0.549
  - **総購買額：**0.730 ／ 0.733，**総購買／日：** 0.686 ／ 0.686

考察

- 総購入額の上位者（Table 6）には極端に大きいものがいる（数千万円を超える）が，CLV 上位者のグループ（Table 7）は必ずしも総購入額の上位者のグループとは一致していない．
- CLV 上位者を特徴づけているのは，購買額よりも，購買回数と直近購買（リーセンシー）である．
  - メディアンで評価すると，Table 7 では Table 6 に比べ，より頻度が高く（30 回以上），よりリーセンシーが小さい（$5 \approx 0.714 \times 7$ 日以内）顧客が上位を構成している．
  - 次の CLV 上位層でも 10 日以内には買い物に来ている．
- 一日当たりの購入額でも一回の店舗での購入額でも Table 7 ではより小さく，グループ間で均等化している．
  - 購入額の回帰式の当てはまりがあまり良くない可能性もある．
  - 購入額とその当てはめ値の相関係数（決定係数）は 0.25 程度．

上位購入者に関する記述統計

Table 6: **上位購入者に関する記述統計（2014 年 12 月から 2017 年 5 月）**

	上位 10%			次の 10%			上位 50%		
	平均	標準偏差	50%点	平均	標準偏差	50%点	平均	標準偏差	50%点
年齢	61.198	15.037	62	58.327	16.166	58	58.143	16.139	58
性別（女性）	0.786	0.41	1	0.798	0.401	1	0.793	0.405	1
購買回数	108.534	92.196	87	71.365	50.37	61	67.528	62.89	51
購買回数/週	0.828	0.696	0.671	0.547	0.381	0.469	0.519	0.476	0.397
直近購買（週）	6.519	16.577	1.143	7.683	17.544	1.714	9.021	19.019	2.143
購買総額（万円）	120.539	250.179	67.394	28.739	4.868	28.052	44.531	132.731	21.586
購買/日（千円）	37.832	247.426	8.494	9.557	23.102	4.577	15.246	125.284	4.618
購買/回（千円）	25.171	109.781	6.426	7.461	20.414	3.577	10.804	56.978	3.597
食品店舗（%）	0.278	0.217	0.24	0.288	0.218	0.252	0.281	0.221	0.244
健康店舗（%）	0.114	0.165	0.044	0.12	0.171	0.044	0.121	0.175	0.043
カフェ（%）	0.121	0.135	0.08	0.108	0.121	0.07	0.111	0.13	0.069
衣料店舗（%）	0.163	0.164	0.116	0.162	0.164	0.115	0.164	0.17	0.113
標本サイズ	4,068			4,078			20,369		

上位 CLV 顧客に関する記述統計

Table 7: 上位 CLV 顧客に関する記述統計（'14 年 12 月から'17 年 5 月）

	上位 10%			次の 10%			上位 50%		
	平均	標準偏差	50%点	平均	標準偏差	50%点	平均	標準偏差	50%点
年齢	61.381	15.183	63	58.223	15.882	58	57.806	16.243	57
性別（女性）	0.795	0.404	1	0.797	0.402	1	0.797	0.402	1
購買回数	138.633	82.401	120	68.624	27.196	66	71.768	60.653	55
購買回数/週	1.060	0.621	0.916	0.529	0.206	0.509	0.553	0.458	0.427
直近購買（週）	7.441	22.046	0.714	9.201	22.752	1.571	10.601	24.069	1.857
購買総額（万円）	76.774	161.919	41.961	36.436	157.773	21.14	38.618	124.447	18.423
購買/日（千円）	6.057	12.487	3.374	5.836	18.789	3.207	6.079	22.56	3.185
購買/回（千円）	4.553	8.657	2.626	4.476	13.431	2.563	4.593	14.347	2.537
食品店舗（%）	0.312	0.219	0.274	0.296	0.213	0.26	0.292	0.22	0.255
健康店舗（%）	0.127	0.174	0.058	0.124	0.174	0.05	0.125	0.177	0.048
カフェ（%）	0.108	0.107	0.078	0.117	0.121	0.079	0.111	0.124	0.072
衣料店舗（%）	0.147	0.146	0.106	0.155	0.149	0.113	0.16	0.162	0.112
生存確率	0.883	0.309	1	0.866	0.318	0.998	0.852	0.327	0.995
CLV（万円）	25.664	14.605	21.186	11.968	1.543	11.784	12.854	10.668	9.57
標本サイズ	4,074			4,073			20,368		

はじめに
データと結果の概要
顧客行動のモデル化と推定
結果の考察
結論と課題
結論と課題
参考文献

結論と課題

結論と課題

- 結論：
  - 購買額／日（短期的利益），総購買額（中期的利益），CLV（長期的利益）をもたらす顧客構成を ROC 曲線で確認
    - CLV は，パレート則ほど上位者が全体に占める割合は高くない
    - 彼らの属性は，（メディアンで）５日前くらいまでに購買行動を行い，これまでも週に一度は買い物を行っていた
    - ただし，支出額は際立って高いわけではない
  - 長期的な視点での利益
    - マーケティング資源を，短期・中期の視点から購買額の上位者，とくに上位 20%までに集中して割くことは「効率的」
    - 顧客がもたらすであろう割引現在価値＝顧客生涯価値にも注目するとき，短期・中期のみの視点からマーケティング資源を過度に配分することは，長期的に価値をもたらす顧客を逃すことになる危険性もある．

結論と課題

- 課題：
  - 購買額の平均の推定を行う回帰分析の当てはまりの向上
  - 購買タイミングの異質性を明示的に考慮したモデル化（例えば，Platzer (2016)）
    - 購買履歴のデータがより本質的に活用できる
  - 店舗選択と来訪・支出の同時決定の考慮

参考文献

[1] 阿部 誠 (2011): 「RFM 指標と顧客生涯価値：階層ベイズモデルを使った非契約型顧客関係管理における消費者行動の分析」,『日本統計学会誌』41(1), 51-81.

[2] Anderson, Cris (2006): *The long tail: Why the future of business is selling less of more*. New York: Hyperion.

[3] Platzer, Michael (2016): "BTYDplus: Probabilistic Models for Assessing and Predicting your Customer Base," available at https://cran.r-project.org/web/packages/BTYDplus/vignettes/BTYDplus-HowTo.pdf