

個人レベルのRFM分析と顧客生涯価値: 新規顧客獲得のための顧客特性への関連付け

阿部 誠

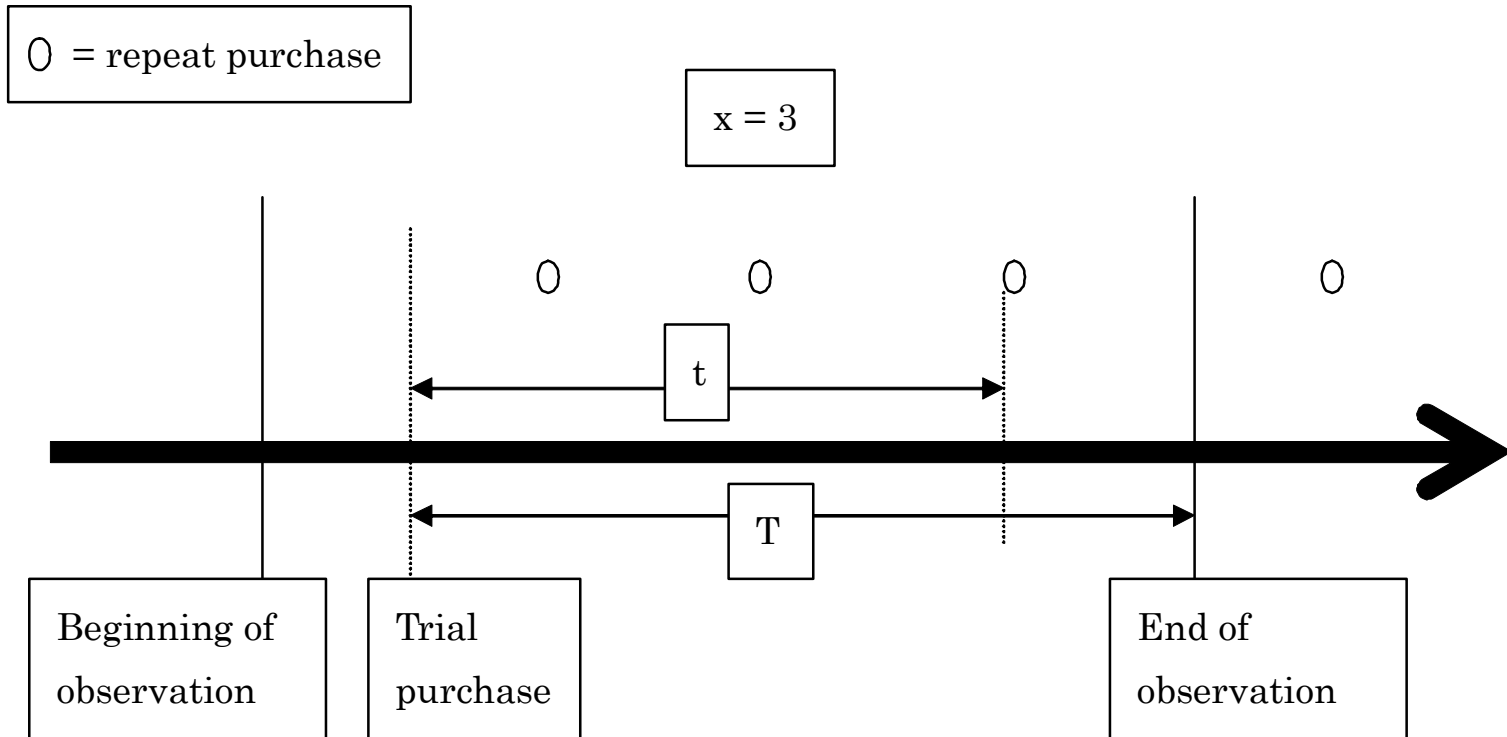
東京大学大学院経済学研究科・経済学部

abe@e.u-tokyo.ac.jp

2006年11月10日 @ 一橋大学COE

1. 動機

顧客生涯価値 (Customer Lifetime Value, CLV) と RFM(recency, frequency, monetary-value)データ



1.1. CLVの計算

- 維持率 (retention rate, r) を知る必要がある.
- “non-contractual” 非契約型の場合、顧客の離脱をどう定義するか?
- Recencyによる判断 (例えば3ヶ月)
 - アドホック、主観的
 - 顧客によって購買間隔が違うことを無視

1.2. スコアリング・モデル

- 維持率 (retention rate) が不要
- 弱点
 - ブラックボックス的なアプローチ
 - Colombo & Jiang (JIM 1999)を参照
 - Fader, Hardie & Lee (JMR 2005)を参照
 - 20-55 と 80-15のルール
(Malthouse & Blattberg JIM 2005)
- 行動モデルの方が有望.

2. 研究の目的

1. 標準的なRFMデータのみを用いて、行動モデルを構築し、非契約型“non-contractual”CRMにおいてCLVを推定.
2. モデルのパラメータを顧客特性に関連付けて、新規顧客獲得に対する知見を得る.

3. モデル

- ・ Pareto/NBD モデル (Schmittlein, Morrison & Colombo, 1987) に基づいた Schmittlein & Peterson (1994) と Fader, Hardie & Lee (2005) の CLV 行動モデルを個人に重点をおいて改良.
- ・ 維持率を個人レベルで推定.
- ・ 経験ベイズではなく階層ベイズの枠組み.

3.1. 顧客行動の仮定

- ・ 購買はポアソン・プロセス
 - 購買はランダムに発生 (rate λ_i)
- ・ 顧客生存期間は指数分布
 - 離脱はランダム (rate μ_i)
- ・ 個人内の購入金額(spending)は対数正規分布
 - 正值のみの分布 (parameter η_i)

$$\log(s_i) \sim N(\log(\eta_i), \omega^2)$$

3.2. 顧客の異質性

- λ_i, μ_i, η_i は多変量対数正規分布 lognormal.
- 顧客特性 d_i との関連付け

$$\begin{bmatrix} \log(\lambda_i) \\ \log(\mu_i) \\ \log(\eta_i) \end{bmatrix} \sim MVN \left(\theta_i = Bd_i, \Gamma_0 = \begin{bmatrix} \sigma_\lambda^2 & \sigma_{\lambda\mu} & \sigma_{\lambda\eta} \\ \sigma_{\mu\lambda} & \sigma_\mu^2 & \sigma_{\mu\eta} \\ \sigma_{\eta\lambda} & \sigma_{\eta\mu} & \sigma_\eta^2 \end{bmatrix} \right)$$

3.3. 既存モデルとの比較

Empirical Bayes Model

Data	Model	Individual Behavior	Heterogeneity Distribution
RF (recency-frequency)	Pareto/NBD (SMC 1987)	Poisson purchase (λ) Random dropout (μ)	$\lambda \sim \text{Gamma}$ $\mu \sim \text{Gamma}$ λ and μ independent
M (monetary-value)	normal-normal (SP 1994)	Normal spending (mean θ)	$\theta \sim \text{Normal}$ θ, λ, μ independent
	gamma-gamma (FHL 2005)	Gamma spending (scale ν)	$\nu \sim \text{Gamma}$ ν, λ, μ independent

Proposed Hierarchical Bayes Model

Data	Model	Individual Behavior	Heterogeneity Distribution
RF (recency-frequency)	poisson/exponential	Poisson purchase (λ) Random dropout (μ)	$\lambda, \mu, \eta \sim \text{MVL}$ λ, μ, η correlated
M (monetary-value)	lognormal-lognormal	Lognormal spending (location η)	

3.4. 顧客生涯価値 (CLV)

$$CLV_i = \int_0^{\infty} v_i(t) s_i(t) d(t) dt = \int_0^{\infty} \lambda_i \eta_i e^{\omega^2/2} e^{-\mu_i t} e^{-\delta t} dt = \frac{\lambda_i \eta_i e^{\omega^2/2}}{\mu_i + \delta}$$

弾力性の分解

$$E_{d_{ik}}^{CLV} = \frac{\partial CLV / CLV}{\partial d_{ik} / d_{ik}} = \left[\frac{\partial CLV}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial d_{ik}} + \frac{\partial CLV}{\partial \mu_i} \frac{\partial \mu_i}{\partial d_{ik}} + \frac{\partial CLV}{\partial \eta_i} \frac{\partial \eta_i}{\partial d_{ik}} \right] \frac{d_{ik}}{CLV}$$
$$= \left[b_{\lambda k} - \frac{b_{\mu k} \mu_i}{\mu_i + \delta} + b_{\eta k} \right] d_{ik} = E f_{d_{ik}}^{CLV} + E r_{d_{ik}}^{CLV} + E m_{d_{ik}}^{CLV}$$

4. 推定

- 経験ベイズ法では個人レベルのパラメータ推定 λ_i, μ_i, η_i が困難.
- 階層ベイズモデルをマルコフチェーン・モンテカルロ (MCMC) 法で推定.
- 非観測(潜在)変数 \Rightarrow Data Augmentation

5. 実証分析

5.1. データ

- 百貨店のフリークエンシー・プログラム (FSP)
- 2000年7月中に入会したメンバーの中から400人をランダムに抽出.
- **購買期間** 7/1/2000~6/31/2001 (52 週！)
 - 前半の26週間は推定用
 - 後半の26週間は検証用
- **3つの顧客特性変数**
 - AGE (/100)
 - FEMALE
 - FOOD = 食料品を購買した来店回数の割合
(店舗へのアクセスの指標)

5.2. モデルの検証

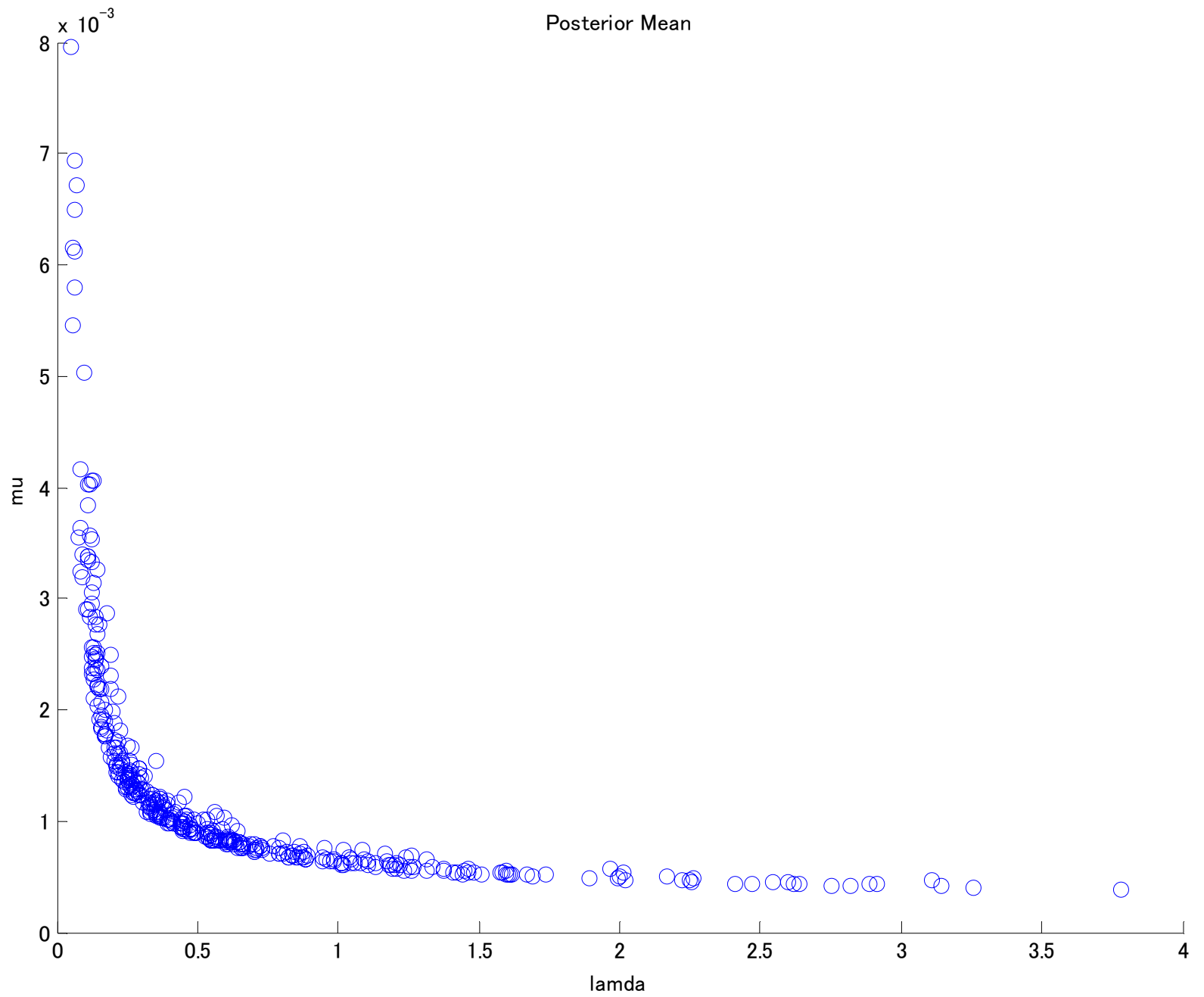
- 15,000 MCMC ステップ
- 後半 5,000 ステップのサンプルから事後分布を構築.
- 推定用、検証用データに対する個人(非集計)レベルでのフィット

		M0	M1	M2	M3
Purchase rate λ	Intercept	-0.886	-1.976	-1.314	-2.036
	Food	---	1.395*	---	1.391*
	Age	---	---	0.533	-0.057
	Female (male=0)	---	---	0.1593	0.100
Dropout Rate μ	Intercept	-7.289	-6.269	-6.638	-6.190
	Food	---	-1.220*	---	-1.066*
	Age	---	---	-0.430	-0.081
	Female (male=0)	---	---	-0.317	-0.238
Spending Parameter η	Intercept	-3.653	-2.883	-3.833	-3.228
	Food	---	-0.987*	---	-1.144*
	Age	---	---	0.274	0.754*
	Female (male=0)	---	---	0.039	0.083
correlation($\log(\lambda)$, $\log(\mu)$)		-0.687*	-0.636*	-0.621*	-0.578*
correlation($\log(\lambda)$, $\log(\eta)$)		-0.285*	-0.153*	-0.298*	-0.158*
correlation($\log(\mu)$, $\log(\eta)$)		0.243	0.131	0.226	0.138
log marginal likelihood		-1947.4	-1944.4	-1932.7	-1925.3

5.3. 推定結果 (相関)

- $\log(\lambda)$ と $\log(\mu)$ の相関係数は -0.58
 - 購買頻度の高い顧客は生存期間が長い.
- この事実は生データからも観測される!
 - 頻度とリーセンサーの相関係数 = -0.41
 - 確率論的にも予期される.
- この相関は、購買頻度が顧客間で大きく異なるほど強くなる傾向にある.
- Pareto/NBDの仮定が犯されている！

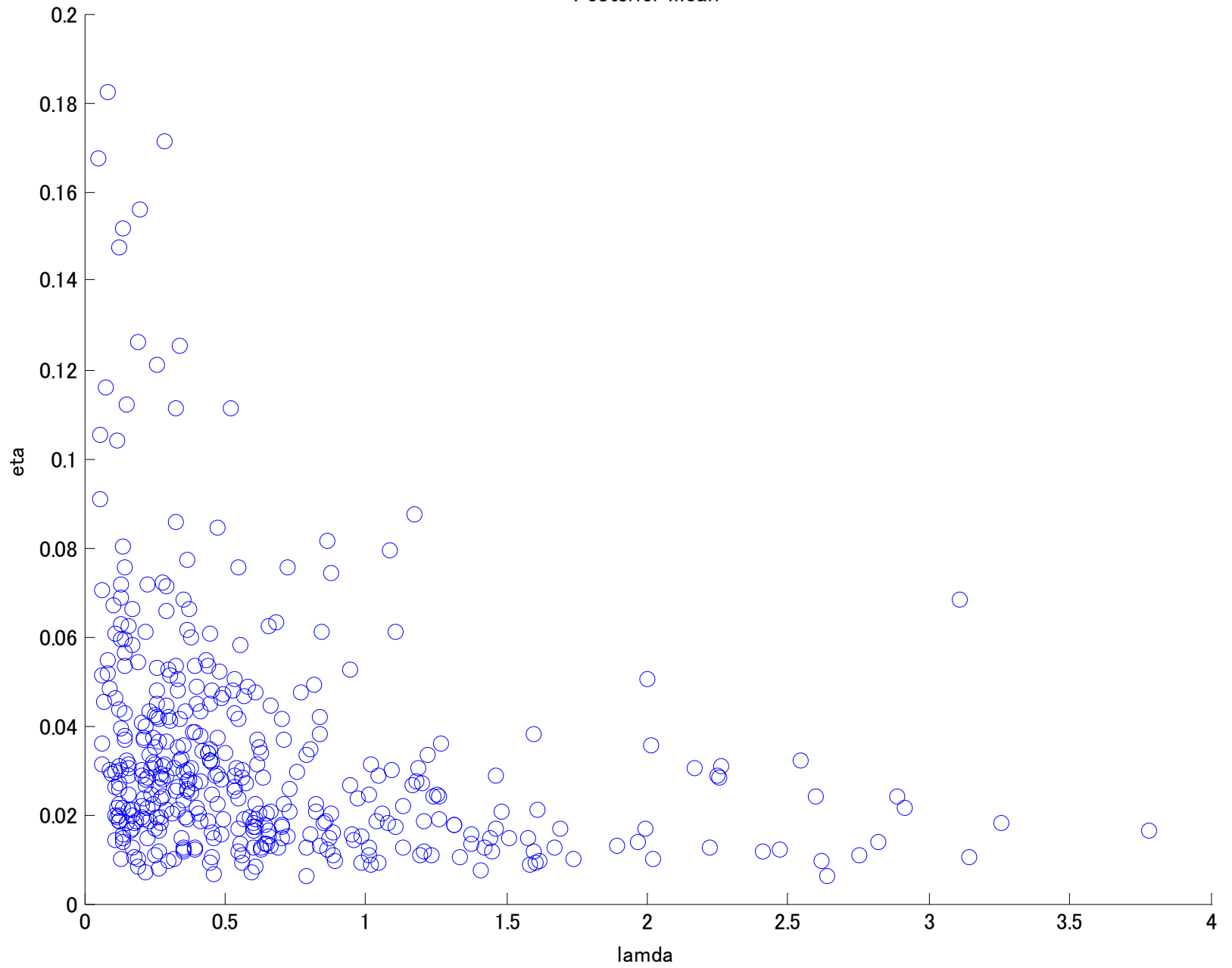
Posterior Mean



5.3. 推定結果 (相関)

- $\log(\lambda)$ と $\log(\eta)$ の相関係数は -0.16.
 - 購買頻度の高い顧客ほど1回あたりの購入金額は低い.
- 購買 (RF) と 金額 (M) は独立でない.
- SP と FHLの仮定が犯されている！

Posterior Mean



		M0	M1	M2	M3
Purchase rate λ	Intercept	-0.886	-1.976	-1.314	-2.036
	Food	---	1.395*	---	1.391*
	Age	---	---	0.533	-0.057
	Female (male=0)	---	---	0.1593	0.100
Dropout Rate μ	Intercept	-7.289	-6.269	-6.638	-6.190
	Food	---	-1.220*	---	-1.066*
	Age	---	---	-0.430	-0.081
	Female (male=0)	---	---	-0.317	-0.238
Spending Parameter η	Intercept	-3.653	-2.883	-3.833	-3.228
	Food	---	-0.987*	---	-1.144*
	Age	---	---	0.274	0.754*
	Female (male=0)	---	---	0.039	0.083
correlation($\log(\lambda)$, $\log(\mu)$)		-0.687*	-0.636*	-0.621*	-0.578*
correlation($\log(\lambda)$, $\log(\eta)$)		-0.285*	-0.153*	-0.298*	-0.158*
correlation($\log(\mu)$, $\log(\eta)$)		0.243	0.131	0.226	0.138
log marginal likelihood		-1947.4	-1944.4	-1932.7	-1925.3

5.4. 推定結果 (係数)

- FOOD

- + for $\log(\lambda)$: 購買頻度が増える
- - for $\log(\mu)$: 離脱率が下がる (生存期間が長くなる)
- - for $\log(\eta)$: (1回あたりの)購入金額が減る

- AGE

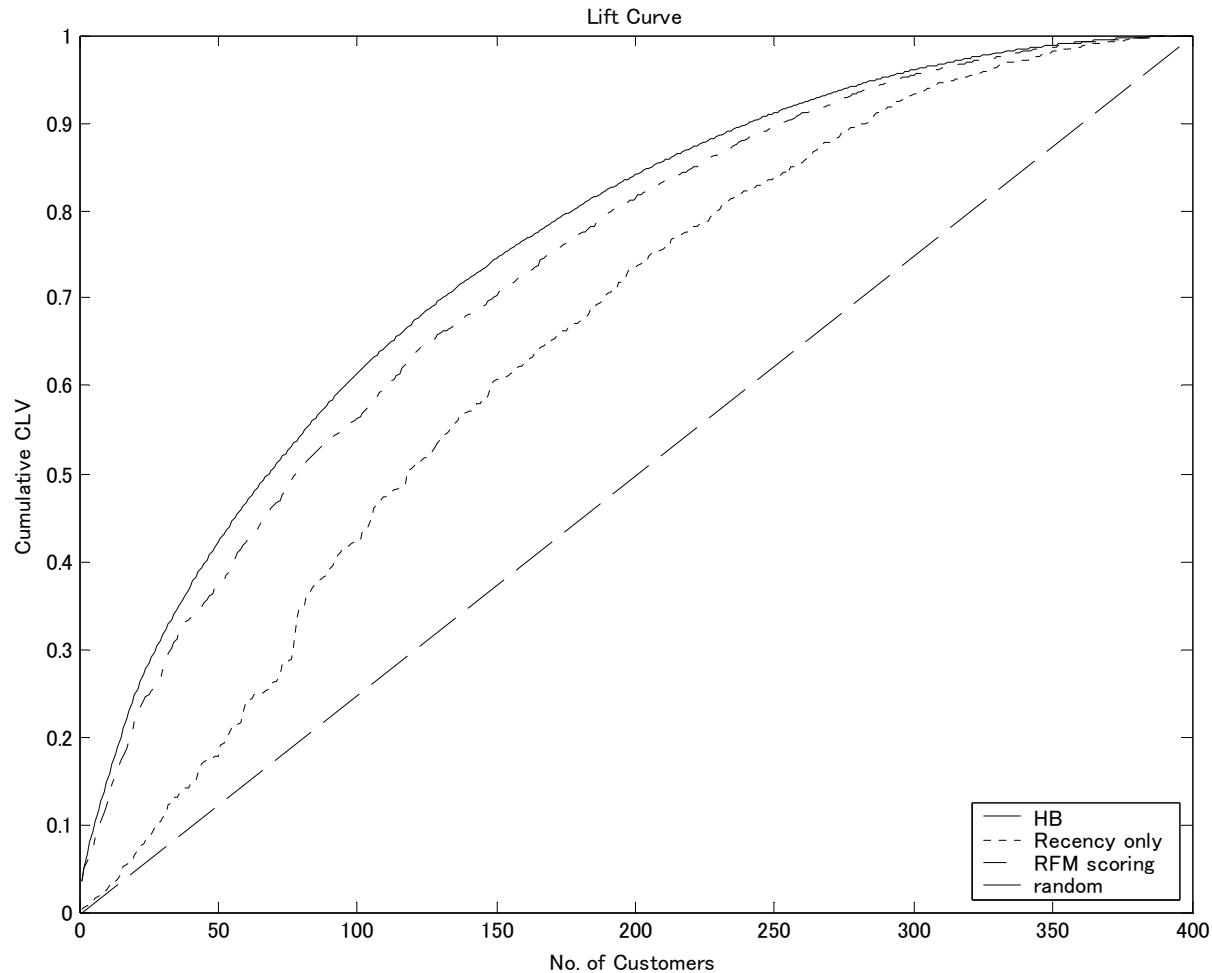
- + for $\log(\eta)$: (1回あたりの)購入金額が増える

5.5. CLV

- 割引率 Discount rate = 15%/年 ($\delta=0.0027$)
- Pareto/NBD モデルでは、 λ_i と μ_i の事後平均、平均生存期間、1年後の維持率の個人レベルでの計算が難しい。

ID	mean(λ)	mean(μ)	mean(η)	Mean Expected lifetime (years)	1 year Retention rate	Probability of being active at the end of calibration	Expected number of transactions in validation period	Expected total spending in val. period ($\times 10^5$ yen)	CLV ($\times 10^5$ yen)
1	0.61	0.00084	0.019	5.6	0.958	0.9997	15.7	0.46	5.3
2	0.97	0.00063	0.024	5.9	0.968	1.0000	25.1	0.93	11.1
3	0.28	0.00125	0.072	5.1	0.938	1.0000	7.1	0.80	8.4
4	0.14	0.00269	0.037	4.1	0.881	0.9771	3.5	0.20	1.8
5	0.86	0.00067	0.018	5.9	0.966	1.0000	22.0	0.63	7.5
6	0.29	0.00148	0.036	4.9	0.928	0.9817	7.3	0.41	4.3
7	1.17	0.00063	0.088	5.9	0.968	0.9997	30.2	4.13	49.6
8	1.60	0.00052	0.012	6.1	0.974	0.9998	41.2	0.75	9.2
9	0.45	0.00104	0.011	5.4	0.950	1.0000	11.6	0.19	2.1
10	0.54	0.00089	0.025	5.6	0.956	0.9960	13.7	0.55	6.2
11	0.57	0.00105	0.019	5.4	0.948	0.9995	14.6	0.43	4.8
12	0.46	0.00092	0.016	5.5	0.954	0.9973	11.7	0.29	3.3
13	0.11	0.00337	0.029	4.2	0.868	0.7430	1.8	0.08	1.1
14	0.72	0.00077	0.015	5.7	0.961	0.9996	18.5	0.44	5.1
15	0.66	0.00076	0.013	5.8	0.962	0.9976	17.1	0.35	4.1
16	1.62	0.00051	0.010	6.1	0.974	0.9987	41.8	0.62	7.7
17	0.29	0.00135	0.171	5.1	0.934	0.9994	7.3	1.95	20.3
18	1.19	0.00058	0.011	6.0	0.971	0.9997	30.8	0.52	6.3
19	0.61	0.00080	0.037	5.7	0.960	0.9989	15.8	0.90	10.4
20	0.40	0.00117	0.053	5.2	0.942	0.9997	10.1	0.85	9.1
ave	0.65	0.00135	0.034	5.3	0.939	0.9749	16.5	0.69	8.0
min	0.05	0.00038	0.006	3.0	0.749	0.6657	1.0	0.03	0.4
max	3.78	0.00795	0.182	6.4	0.981	1.0000	97.9	8.59	107.6

ゲイン・チャートの比較



- スコアリング・モデルも優良顧客のランキングには有用

5.6. 弾力性 (I)

- CLVの弾力性を、頻度、維持、金額の3要因に分解

$$E_{d_{ik}}^{CLV} = E_{f_{d_{ik}}}^{CLV} + E_{d_{d_{ik}}}^{CLV} + E_{s_{d_{ik}}}^{CLV} = \left[b_{\lambda k} - \frac{b_{\mu k} \mu_i}{\mu_i + \delta} + b_{\eta k} \right] d_{ik}$$

	FOOD	AGE	FEMALE
Total	0.403	0.383	0.233
<i>frequency: E_f^{CLV}</i>	1.095	-0.030	0.093
<i>dropout: E_d^{CLV}</i>	0.208	0.016	0.063
<i>spending: E_s^{CLV}</i>	-0.901	0.397	0.077

5.6. 弾力性 (II)

- CLVの弾力性を、頻度、維持、金額の3要因に分解

$$E_{d_{ik}}^{CLV} = Ef_{d_{ik}}^{CLV} + Ed_{d_{ik}}^{CLV} + Es_{d_{ik}}^{CLV} = \left[b_{\lambda k} - \frac{b_{\mu k} \mu_i}{\mu_i + \delta} + b_{\eta k} \right] d_{ik}$$

- 3要因はCLVに対してお互いに逆の影響をもたらす可能性もある

5.6. 弾力性 (III)

- CLVの弾力性を、頻度、維持、金額の3要因に分解

$$E_{d_{ik}}^{CLV} = E f_{d_{ik}}^{CLV} + E d_{d_{ik}}^{CLV} + E s_{d_{ik}}^{CLV} = \left[b_{\lambda k} - \frac{b_{\mu k} \mu_i}{\mu_i + \delta} + b_{\eta k} \right] d_{ik}$$

- 3要因はCLVに対してお互いに逆の影響をもたらす可能性もある
- 非線形なモデルパラメータでは、推定誤差（不確実性）を考慮することが大切
⇒ 意思決定や最適化において

パラメータの不確実性を考慮した場合

	FOOD	AGE	FEMALE
Total	0.403	0.383	0.233
<i>frequency: E_f^{CLV}</i>	1.095	-0.030	0.093
<i>dropout: E_d^{CLV}</i>	0.208	0.016	0.063
<i>spending: E_s^{CLV}</i>	-0.901	0.397	0.077

パラメータの不確実性を無視した場合

	FOOD	AGE	FEMALE
Total	0.474	0.381	0.244
<i>frequency: E_f^{CLV}</i>	1.095	-0.030	0.093
<i>dropout: E_d^{CLV}</i>	0.279	0.014	0.074
<i>spending: E_s^{CLV}</i>	-0.901	0.397	0.077

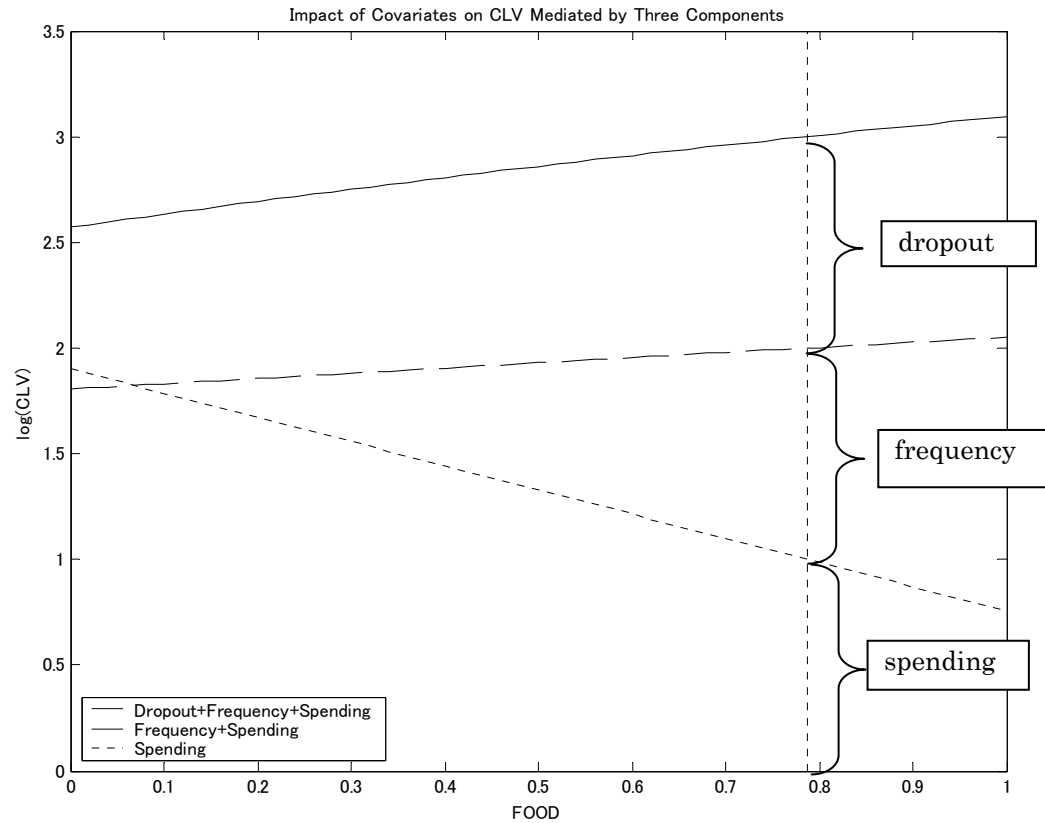
5.7. Elasticity (IV)

- 違った視点からのCLVの分解

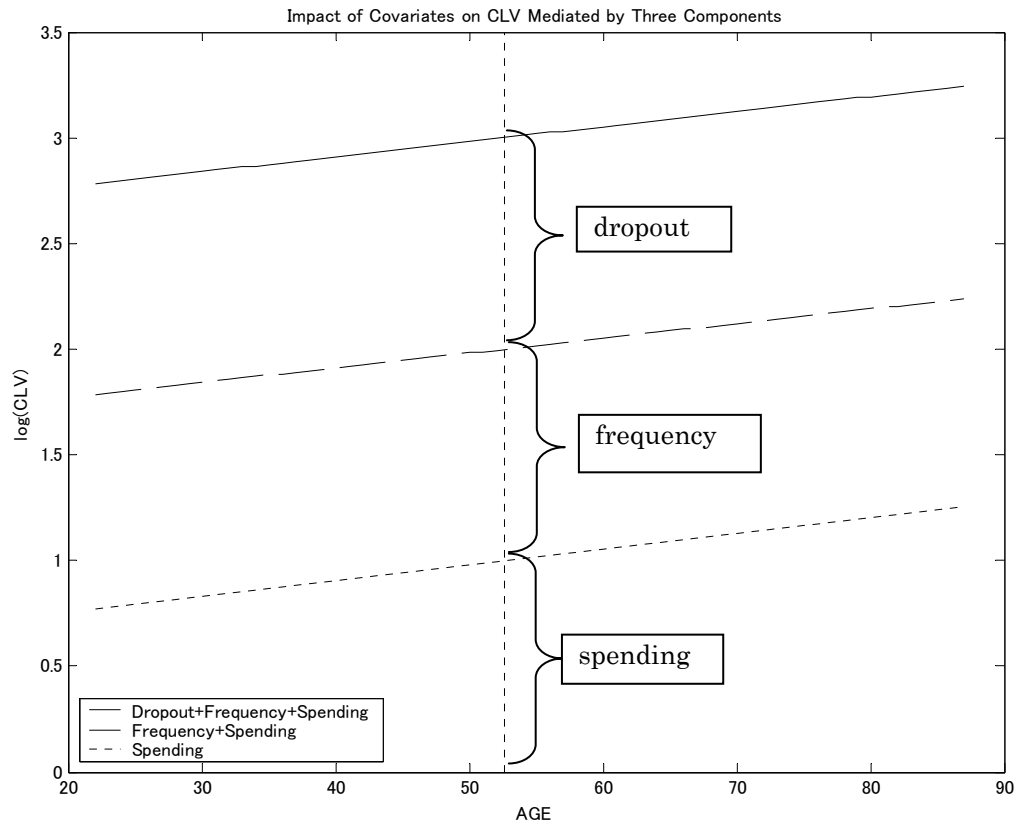
$$CLV_i = \frac{\lambda_i \eta_i e^{\omega^2/2}}{\mu_i + \delta}$$

- 対数をとって和の形に変換し、顧客特性変数を横軸にグラフをプロット.
- 3要因は $\log(\text{CLV})$ に対してお互いに逆の影響をもたらす可能性がある.

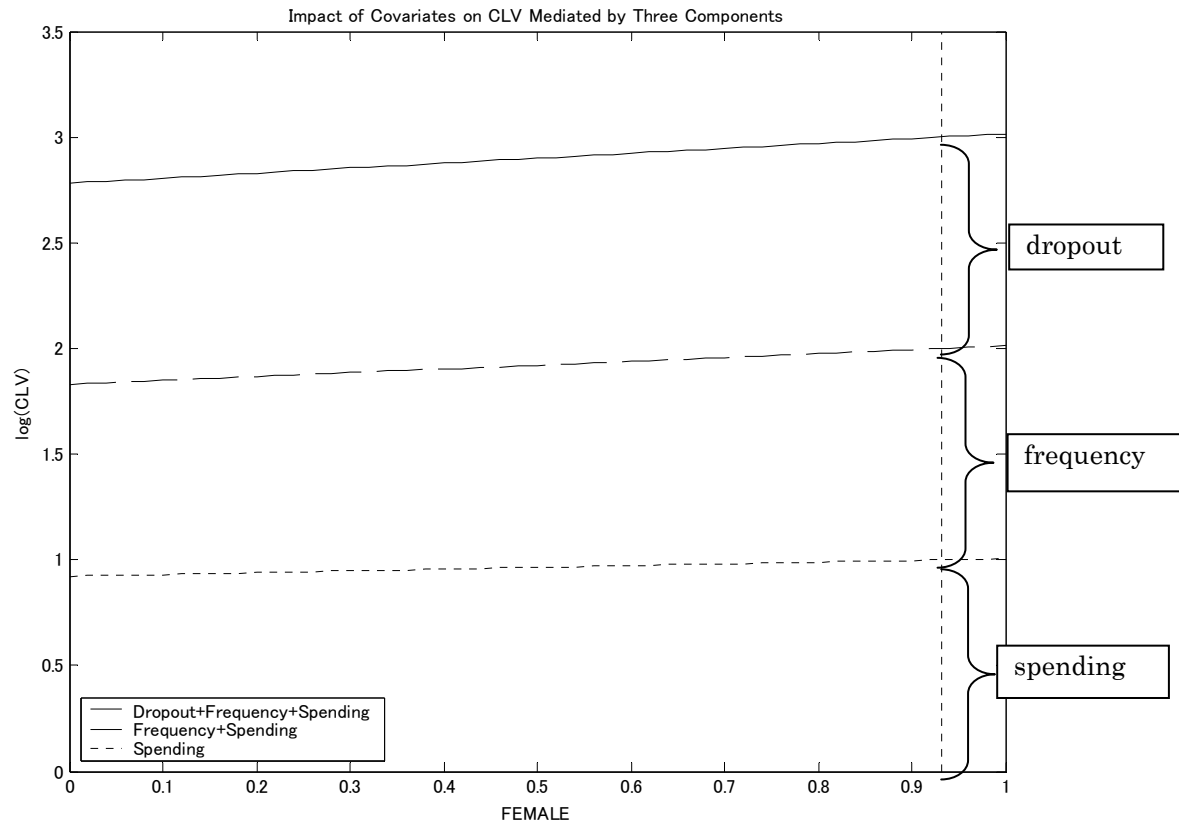
FOOD 変数



AGE 変数



FEMALE 変数



5.8. 新規顧客獲得

- CLVが高いのは
 1. 年配の顧客
 2. 食料品の購買者

6. 結論

1. 手法的な新規性
2. マーケティングに関する本質的な知見
3. マネジエリアル・インプリケーション

6.1. 手法における新規性

1. RFMデータに対応した個人レベルの行動モデルを階層ベイズの枠組みに基づいて構築.
 - 経験ベイズ法に基づいた既存モデルよりもシンプルで、より複雑なモデル拡張が可能.
2. MCMC法にデータ補完を組み込むことによって、観測できない変数の個人レベルでの推定が可能.

6.2.マーケティングに関する本質的な知見

1. Pareto/NBD モデルの仮定である購買頻度と離脱プロセスの独立性が、満たされないことがある。(負の相関)
2. Pareto/NBD モデルの仮定である購買と金額の独立性が、満たされないことがある。(負の相関)
3. CLVの弾力性を頻度、維持、金額の3要因に分解. 各要因はお互いにCLVに逆の影響をもたらすこともある.
4. パラメータの不確実性を考慮する重要性

6.3. マネジリアル・インプリケーション

1. CRMにおける顧客別潜在変数の有用性
(Pareto/NBDモデルでは計算が困難)
2. 新規顧客獲得に関する知見
3. パラメータの不確実性を考慮する必要性.
(意思決定や最適化において)