顧客情報を考慮した
エリア・セグメンテーションに関する研究

東京理科大学大学院
工学研究科　経営工学専攻
修士2年　石田　佳之

目次

- 研究背景
- 研究目的
- データ概要
- 分析
  - モデル1概要
  - モデル2概要
  - モデル1結果・考察
  - モデル2結果・考察
- まとめと今後の課題
- 参考文献

エリア・マーケティングの重要性

- **エリア・マーケティング**
  - ✓ 全国一律ではなく，エリア（地域）ごとのニーズ特性や競合状況に応じて商品・サービス，価格，プロモーション，流通などのマーケティングの方法を変更すること.

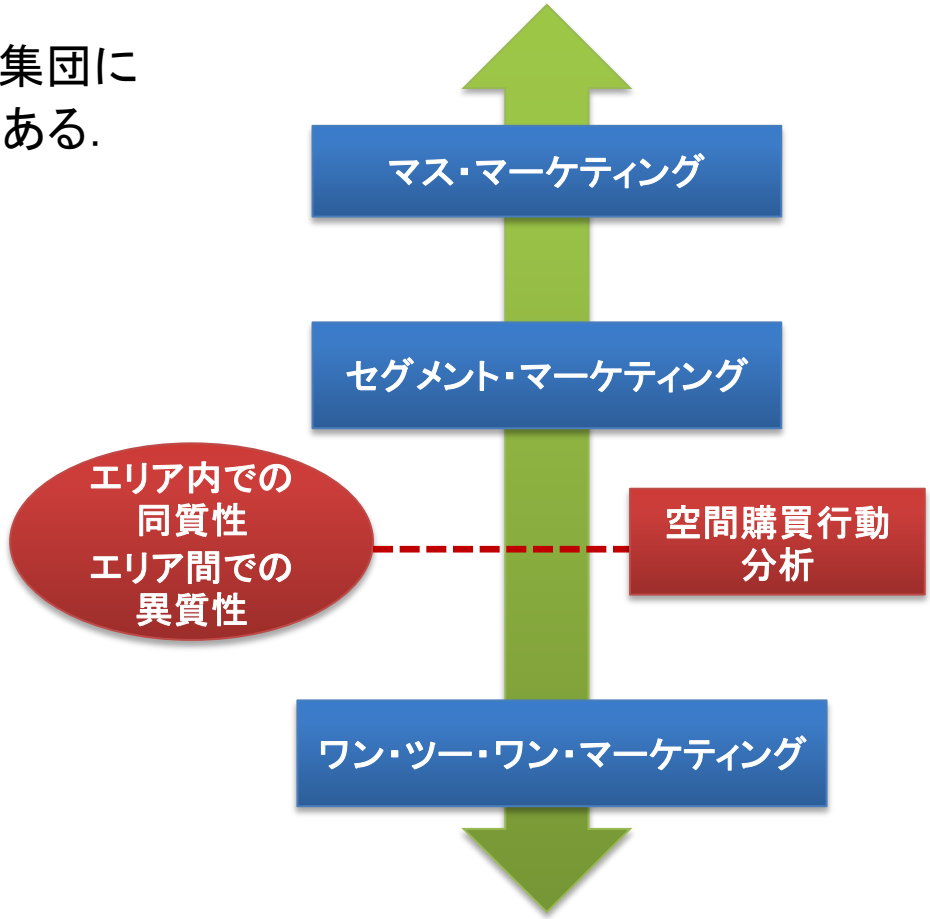
- 店舗独自の地域密着型マーケティングが必須条件[7]
  - 消費者ニーズの多様化，メーカー商品の品質拮抗
  - 市場・商圏・顧客情報などの各種データには必ず地域差が存在

- 企業は店舗によって，異なる商圏の需要に対応するために品揃えや価格戦略を変更している.

エリア・セグメンテーション

- マーケット・セグメンテーション
  - 対象をいくつかの集団に分類すること.
  - 市場には異質な集団があるため，特定集団に対してマーケット活動を行う方が効率である.

- エリア・セグメンテーション
  - 類似性が高いエリア同士でグループを作り，エリア内での同質性，エリア間での異質性を把握すること.
  - セグメンテーション要件のうち，特定可能性・到達可能性を満たす.
  - GIS（Geographic Information System）の活用の際，複数情報の集約に重要.



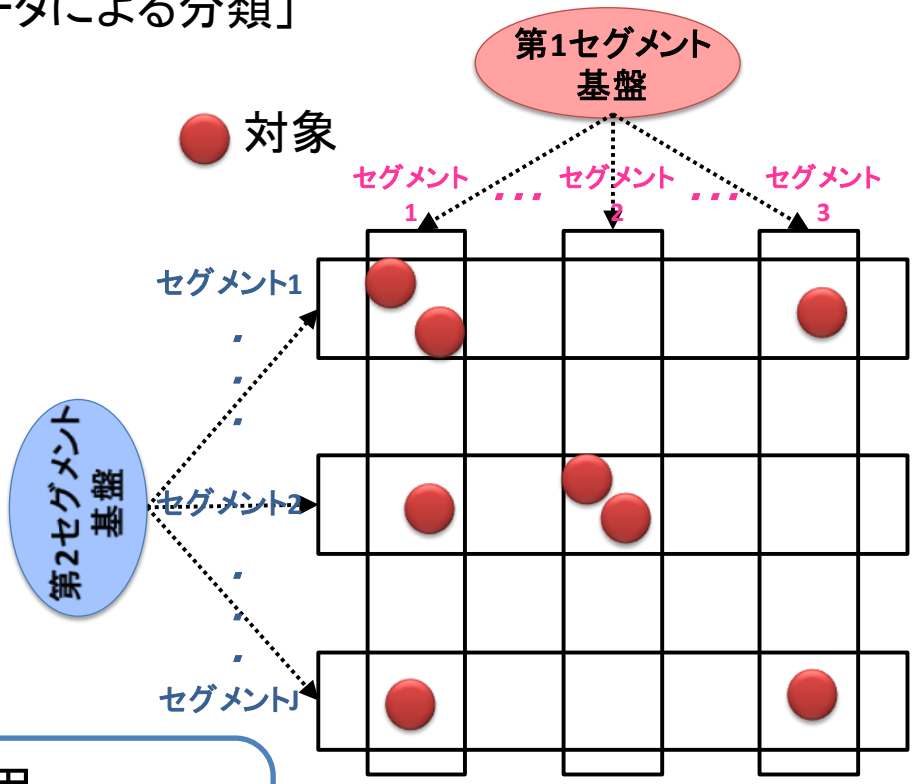
ID付きPOSデータの登場

- 従来のエリア・セグメンテーションの問題点
  - 従来はセンサスデータのみをもとに，分類を行っていた．
  - センサス・データでは顧客の購買反応の違いが測定できないため，店舗から見た顧客・市場の評価が反映できない
  - 顧客維持だけでなく顧客獲得も考えると，センサス・データのみでは不十分
- ID付きPOSデータの登場
  - 顧客の購買履歴情報から，顧客情報を容易に入手可能に．
  - その結果，顧客・市場の評価を反映させた，セグメンテーションが可能に．

ジョイント・セグメンテーション

- **ジョイント・セグメンテーション**
  - 複数のセグメント軸を同時に利用してセグメントを形成する方法[2].
  - EX)「顧客属性による分類」×「購買データによる分類」
  - 既存研究
    - V.Ramaswamyら[2]
    - 里村ら [8]，坂巻[7]
    - Rick L. Andrewsら[10]
  - 潜在クラスモデルの1つ

    観測上は存在しないが，いくつかの潜在的なグループが存在すると仮定し，事後的にセグメントを形成するモデル.



1. GISを用いる際の情報の集約に有用.
2. セグメント基盤に顧客情報を用いることで，購買行動ベースでのエリア分類が可能

里村[1]の先行研究

- ジョイント・セグメンテーションの商圏分析への活用
  - 里村[1]，“商圏分析のためのエリア・セグメンテーション”
    - 地方百貨店の1つの店舗を対象に，ジョイント・セグメンテーションを用いてエリア・セグメントを形成．
    - 「顧客属性」×「購買データ」によるエリアの分類．
  - 問題点
    1. 1つの店舗を扱ったモデルであり，競合他社や自社他店舗が考慮されていない．
    2. マーケティングにおける商圏の定義[9]
       - 「小売施設の顧客が住む地域」
       - 「来店客の70%が居住し，しかもその人々が来店する際の平均旅行時間（時間）が最小となるような居住区域の集合」

    ⇔ 各エリアの顧客規模が考慮されておらず，同一と仮定したセグメントを形成している．

研究目的

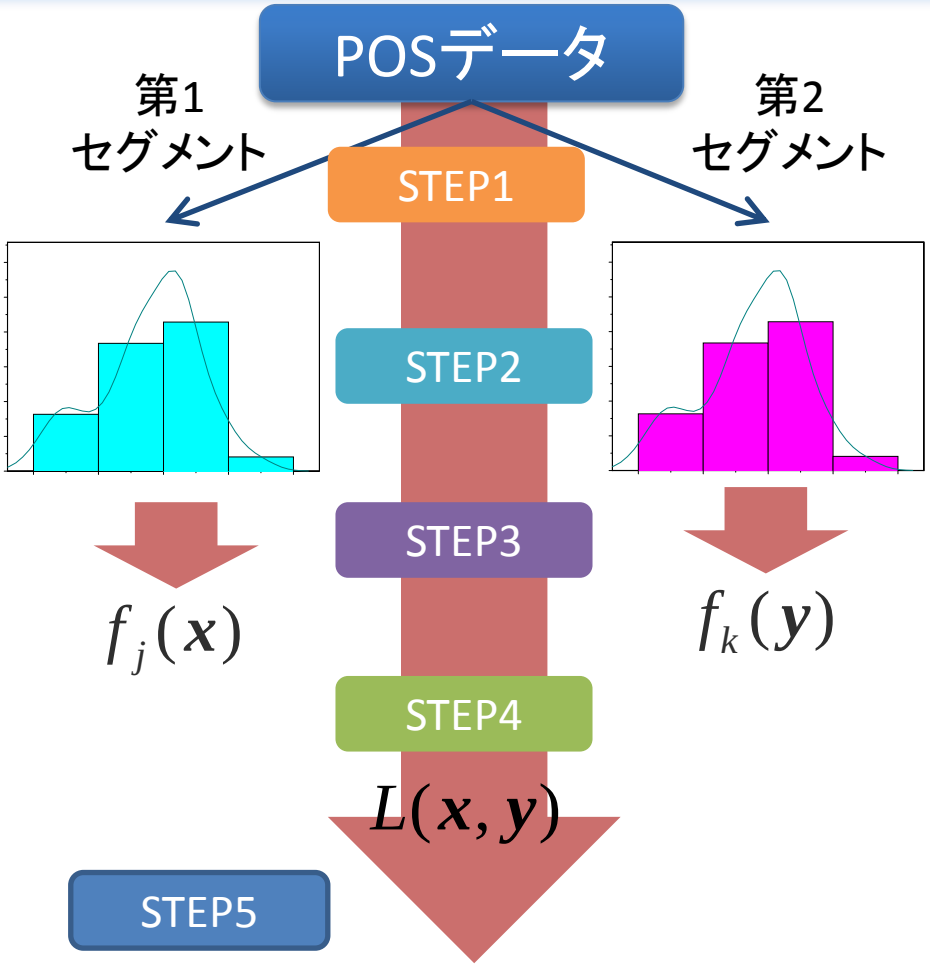
- 研究背景
  - エリア・マーケティングは流通業界では必須条件であり，顧客に効率的にアプローチするためにエリア・セグメンテーションがある．
  - 従来のエリア・セグメンテーションでは，センサスデータによる分類を行ってきたが，顧客情報が組み込まれてこなかった．

    ⇔　ID付きPOSデータの登場で，市場・顧客の評価を反映したセグメンテーションが容易に可能になった．
  - 里村のモデルでは，①競合他社や自社他店舗を考慮していない．②商圏を知る際に重要な各エリアの顧客規模を考慮していない．
- 研究目的
  - 都内百貨店の複数店舗を対象にジョイント・セグメンテーションモデルを用いて，顧客の購買情報をもとにエリア・セグメントを作成し，都内における消費の分布を確認する．
  - 本研究の対象百貨店での「既存店舗の評価」という観点から，自社他店舗を考慮する．
  - 商圏の設定に必要なエリアごとの顧客規模を考慮する．

データ概要

- 2009年度データ解析コンペティション提供データ
- 都内百貨店ID付きPOSデータ
- 店舗：有楽町，池袋，渋谷の3店舗
- データ期間：2008年4月1日～2009年3月31日
- 顧客数：約55万人
- データ属性
  - 顧客属性：
    - 顧客番号，性別，購入時年齢，住所コード（上5ケタ）
  - 商品属性：
    - 売上数量，売上高，清算番号，プロパーバーゲン区分，売上解約区分，行番号，基本アイテムコード
  - 店舗・時間属性：
    - 店コード，ブロックコード，清算時刻，売上日

ジョイント・セグメンテーションの分析ステップ



1. 各セグメントの解釈
2. セグメントの構成比
3. 各エリアのセグメントへの所属確率

- **STEP1**
  - セグメント基盤の決定
- **STEP2**
  - 各セグメント基盤に分布を仮定
- **STEP3・STEP4**
  - 各基盤の尤度，全体の尤度を計算
- **STEP5**
  - パラメータの推定
    （EMアルゴリズム）

既存研究と本研究の比較

- モデル1では，顧客規模を考慮(Joint Approach).
- モデル2では，1つのセグメント基盤で2つの変数を用いて，顧客の店舗選択の影響を考慮(Joint Approach + Combined Approach).

モデル	第1セグメント基盤		第2セグメント基盤	
	変数	分布	変数	分布
里村[1]のモデル ※APPENDIX[2]参照	各エリアの顧客属性ごとの顧客数	**多項分布**	各エリアのカテゴリごとの平均購買金額	正規分布
モデル1	各エリアの顧客属性ごとの顧客数	**正規分布**	各エリアのカテゴリごとの平均購買金額	正規分布
モデル2	・各エリアの顧客属性ごとの顧客数 ・各エリアの店舗毎の顧客数	・**正規分布** ・**多項分布**	各エリアのカテゴリごとの平均購買金額	正規分布

記号の定義

- 添字の定義

j : 第1セグメント基盤の第j番目のセグメント$(j = 1,..J)$

k : 第2セグメント基盤の第k番目のセグメント$(k = 1,..K)$

g : エリア$g(g = 1,..G)$

i : カテゴリ$i(i = 1,..I)$

d : 顧客属性$d(d = 1,..D)$

- 変数の定義

x_{gd} : エリアgの顧客属性 dの顧客数

y_{gi} : エリアgの商品分類 iの平均購買金額

表1 : 使用データの一例

性別	年齢
女性	20代前半
女性	20代後半
女性	30代前半
男性	20代前半
男性	20代後半

カテゴリ
アクセサリー・時計
インテリア・電化製品
レディース
メンズ
スポーツ
化粧品

住所コード	住所名称
13101	東京都　千代田区
13102	東京都　中央区
13103	東京都　港区
13104	東京都　新宿区
13105	東京都　文京区
13106	東京都　台東区

モデル1（STEP2・STEP3）

1. 各基盤に分布を仮定
   - 第1セグメント基盤：平均β_{jd}，分散σ^2_{jd}の独立した正規分布
   - 第2セグメント基盤：平均β_{ki}，分散σ^2_{ki}の独立した正規分布
2. 各基盤の尤度の計算
   - 第1セグメント基盤

$$f_j(\boldsymbol{x_g}) = f_j(x_{g1},\ldots,x_{gD})$$

$$= \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\sigma^2_{jd}}} \exp\left(-\frac{1}{2\sigma^2_{jd}} \left(x_{gd} - \beta_{jd} \right)^2 \right) \qquad (1)$$

   - 第2セグメント基盤

$$f_k(\boldsymbol{y_g}) = f_k(y_{g1},\ldots,y_{gI})$$

$$= \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma^2_{ki}}} \exp\left(-\frac{1}{2\sigma^2_{ki}} \left(y_{gi} - \beta_{ki} \right)^2 \right) \qquad (2)$$

$f_j(x_g)$ ：エリアgの第1セグメント基盤のセグメントjの尤度

$f_k(y_g)$ ：エリアgの第2セグメント基盤のセグメントkの尤度

モデル1（STEP4・STEP5）

- エリアg全体の尤度

$$L_g(\boldsymbol{x_g}, \boldsymbol{y_g}) = \sum_{j=1}^{J}\sum_{k=1}^{K} \phi_{jk} \left\{ f_j(\boldsymbol{x_g}) \cdot f_k(\boldsymbol{y_g}) \right\} \quad (3)$$

ただし，$0 < \phi_{jk} < 1, \quad \sum_{j=1}^{J}\sum_{k=1}^{K} \phi_{jk} = 1 \quad (4)$ ϕ_{jk}：jkセグメントへの所属確率

- データ全体の尤度

$$L = \prod_{g=1}^{G} L_g(\boldsymbol{x_g}, \boldsymbol{y_g}) = \prod_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K} \phi_{jk} \left\{ f_j(\boldsymbol{x_g}) \cdot f_k(\boldsymbol{y_g}) \right\} \quad (5)$$

パラメータ
ϕ_{jk}, β_{jd}, σ^2_{jd}, β_{ki}, σ^2_{ki}

- 尤度を最大にするパラメータを推定．
- 推定には，潜在クラスモデルで一般的に利用される**EMアルゴリズム**を利用．（APPENDIX[1]参照）

- パラメータの推定後，エリアgが所属するセグメントは事後確率が最大になるセグメントとして決定．

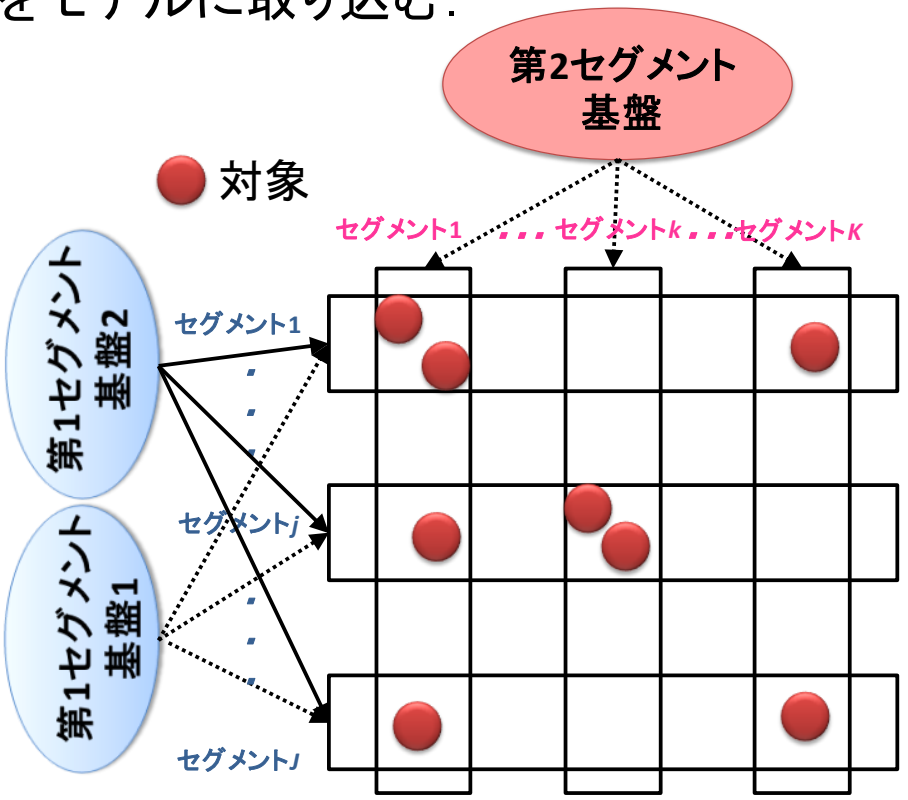
$$\Pr(g \in j,k) = P(j,k \mid g) = \frac{\phi_{jk} \left\{ f_j(\boldsymbol{x_g}) \cdot f_k(\boldsymbol{y_g}) \right\}}{\sum_{j=1}^{J}\sum_{k=1}^{K} \phi_{jk} \left\{ f_j(\boldsymbol{x_g}) \cdot f_k(\boldsymbol{y_g}) \right\}} \quad (6)$$

モデル2（STEP1）

- 店舗間の関係を考慮するために，モデル1にさらに他店舗顧客数の変数をモデル2に取り込む．
- しかし，3変数以上を扱うと，結果が3次元になり，セグメントサイズが過大に．
  → 1つのセグメント基盤に2つの変数をモデルに取り込む．
- 第1セグメント基盤の決定
  1. 渋谷店の顧客属性ごとの顧客数
  2. 各店舗の顧客数

表：各店舗の顧客数の変数

各店舗の顧客数
有楽町店のみを利用する顧客数
池袋店のみを利用する顧客数
渋谷店のみを利用する顧客数
有楽町店と池袋店を利用する顧客数
有楽町店と渋谷店を利用する顧客数
渋谷店と池袋店を利用する顧客数
全3店舗を利用する顧客数



図：モデル2のイメージ

モデル2（STEP2・STEP3）

- 各軸の尤度の計算
  - 第1セグメント基盤

1. 顧客属性ごとの顧客数の分布
   → 独立した正規分布
2. 各店舗の顧客数の分布
   → 多項分布

1. $f_j(\boldsymbol{x_g}) = f_j(x_{g1},\ldots,x_{gD})$

$$= \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\sigma_{jd}^2}} \exp\left(-\frac{1}{2\sigma_{jd}^2}\left(x_{gd} - \beta_{jd}\right)^2 \right) \qquad (7)$$

2. $g_j(\boldsymbol{w_g}) = g_j(w_{g1},\ldots,w_{gC})$

$$= \left(\sum_{c=1}^{C} w_{gc} \right)! \prod_{c=1}^{C} \left(\frac{\theta_{jc}^{w_{gc}}}{w_{gc}!} \right) \qquad (8)$$

c：変数cの顧客数$(c = 1,2,\ldots, C)$

θ_{jc}：セグメントjの変数cの多項分布パラメータ

w_{gc}：各店舗での顧客数

$$F_j(\boldsymbol{x_g}, \boldsymbol{w_g}) = f_j(\boldsymbol{x_g}) \cdot g_j(\boldsymbol{w_g}) \qquad (9)$$

$g_j(\boldsymbol{w_g})$：エリアgの第1セグメント基盤2のセグメントjの多項分布の尤度

$F_j(\boldsymbol{x_g}, \boldsymbol{w_g})$：エリア$g$の第1セグメント基盤のセグメント$j$の尤度

モデル2（STEP4・STEP5）

- エリアg全体の尤度

$$L_g(\boldsymbol{x_g}, \boldsymbol{y_g}, \boldsymbol{w_g}) = \sum_{j=1}^{J}\sum_{k=1}^{K}\phi_{jk}\left(F_j(\boldsymbol{x_g}, \boldsymbol{w_g}) \cdot f_k(\boldsymbol{y_g})\right) = \sum_{j=1}^{J}\sum_{k=1}^{K}\phi_{jk}\left(\left(f_j(\boldsymbol{x_g}) \cdot g_j(\boldsymbol{w_g})\right) \cdot f_k(\boldsymbol{y_g})\right) \quad (10)$$

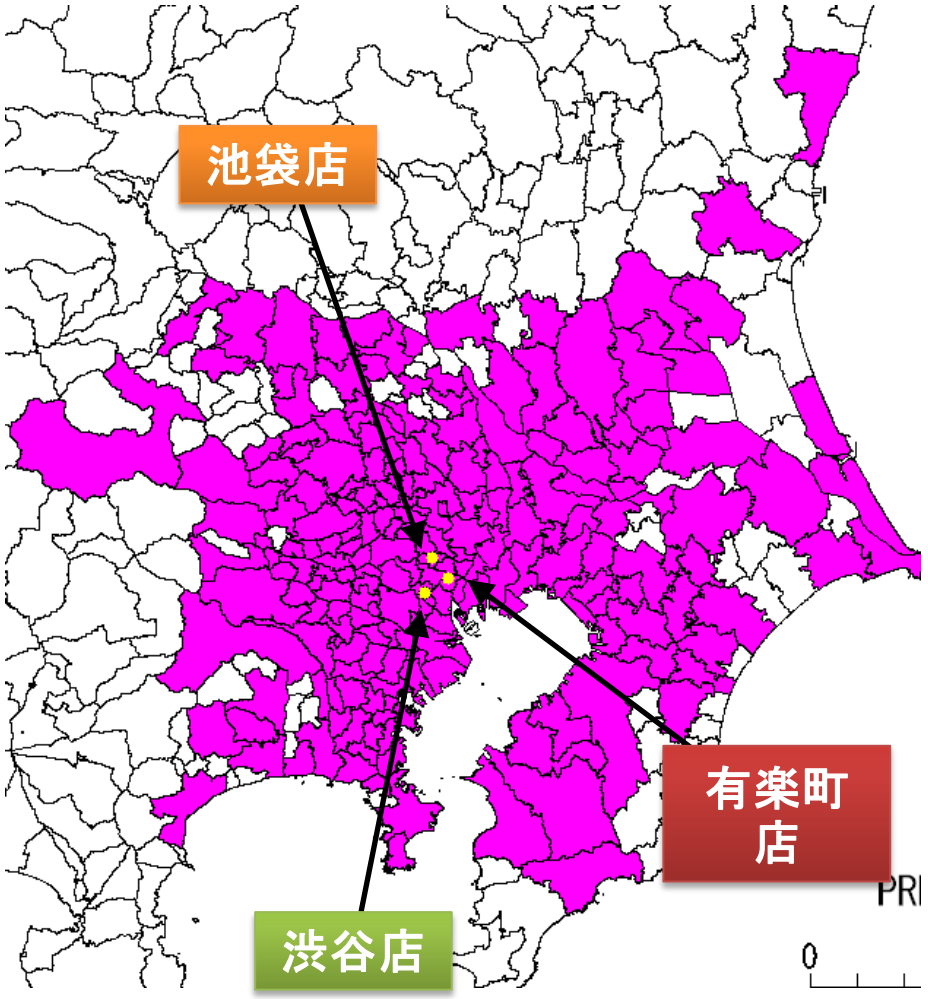
- データ全体の尤度

$$L = \prod_{g=1}^{G} L_g(\boldsymbol{x_g}, \boldsymbol{y_g}, \boldsymbol{w_g}) = \prod_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K}\phi_{jk}\left(\left(f_j(\boldsymbol{x_g}) \cdot g_j(\boldsymbol{w_g})\right) \cdot f_k(\boldsymbol{y_g})\right) \quad (11)$$

- パラメータ ϕ_{jk}, θ_{jc}, β_{jd}, σ^2_{jd}, β_{ki}, σ^2_{ki}の推定
  - 同様に，尤度（式（11））が最大になるように，EMアルゴリズムを用いて推定する.

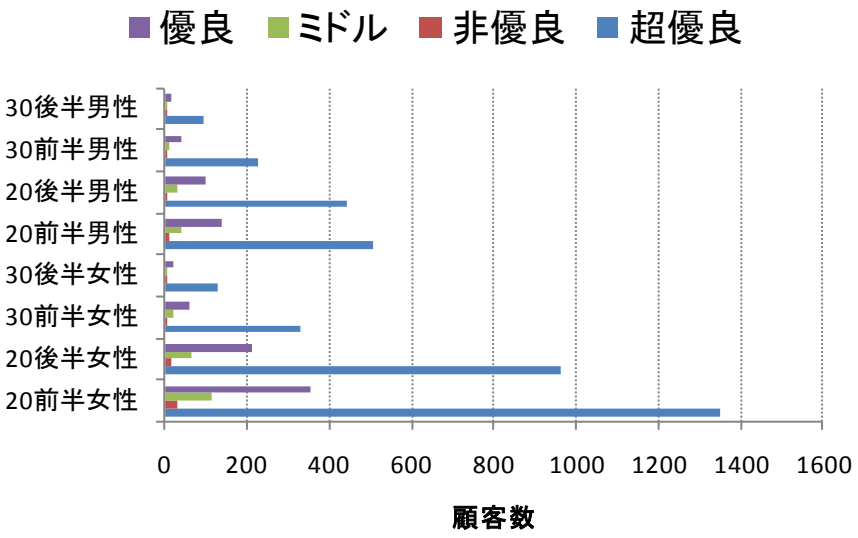
分析対象エリア

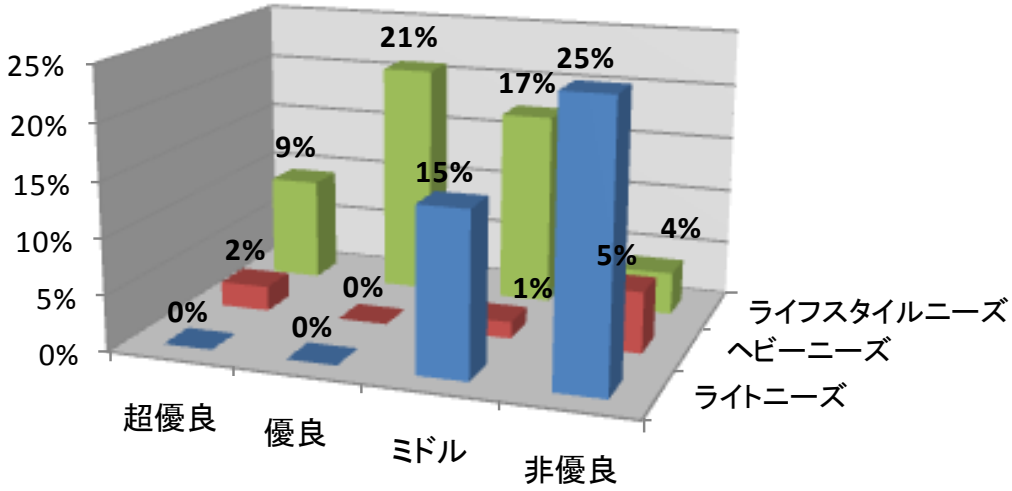
- 関東圏のうち，年間100人以上の顧客がいる207エリア.
- 住所区分は郵便番号で分類.
- 本研究では自社の3店舗のみを考え，競合店舗やその他自社店舗は考慮しない.
  - 渋谷店（主分析対象）
  - 有楽町店
  - 池袋店

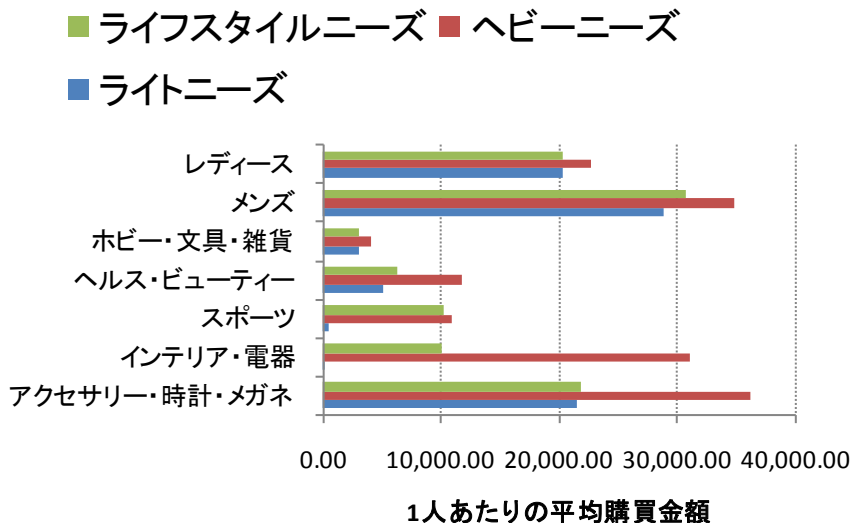


モデル1：分析結果（渋谷店）

- BICの値により各セグメント基盤のセグメント数を選択
  - 顧客デモグラフィックス：4セグメント
  - 購買金額：3セグメント
  - BIC：42185.21

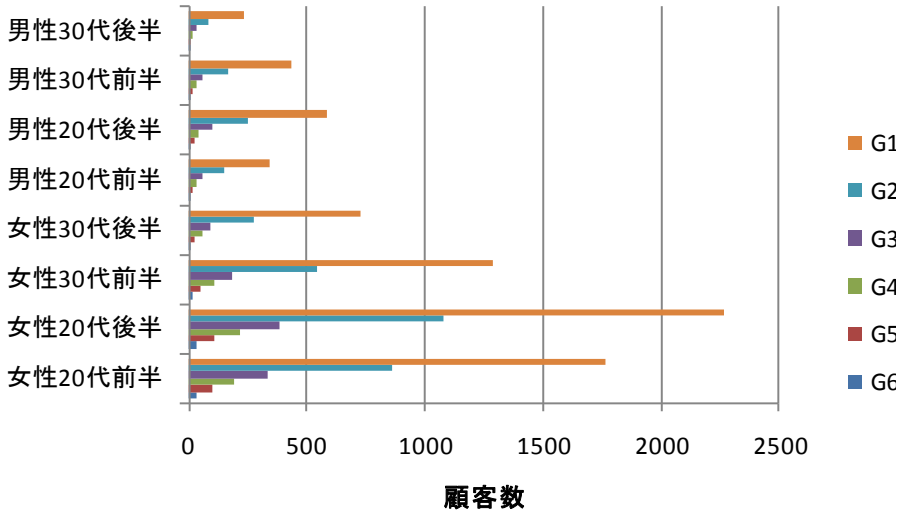


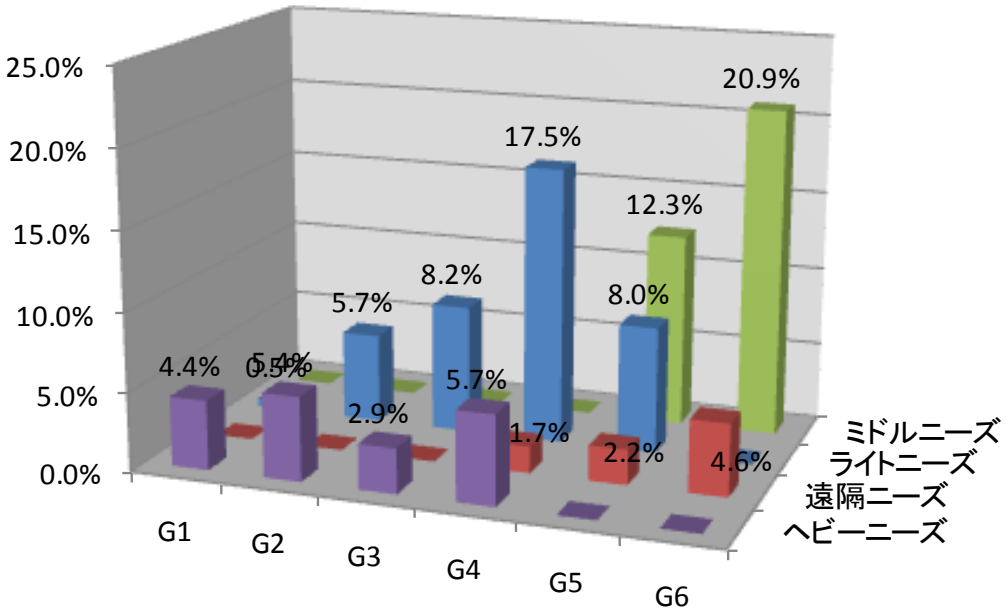


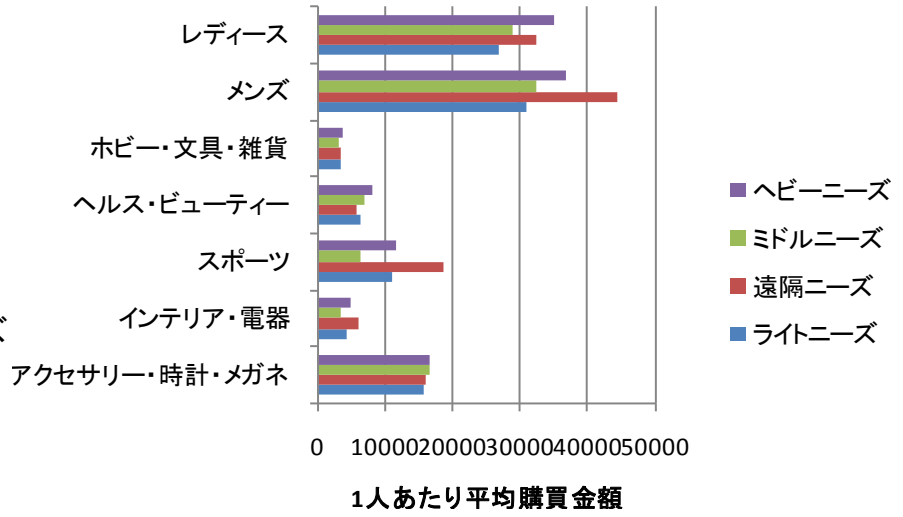


モデル1:分析結果(有楽町店)

- BICの値により各セグメント基盤のセグメント数を選択
  - 顧客デモグラフィックス:6セグメント
  - 購買金額:4セグメント
  - BIC:41932.70

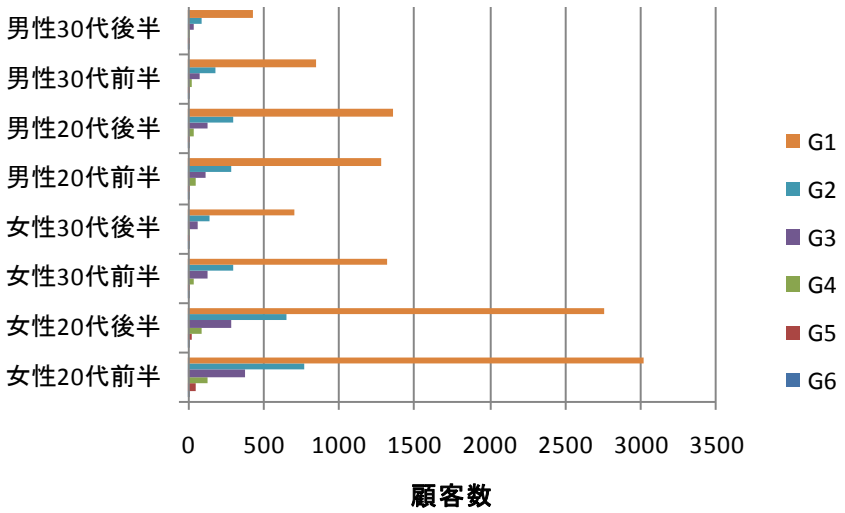


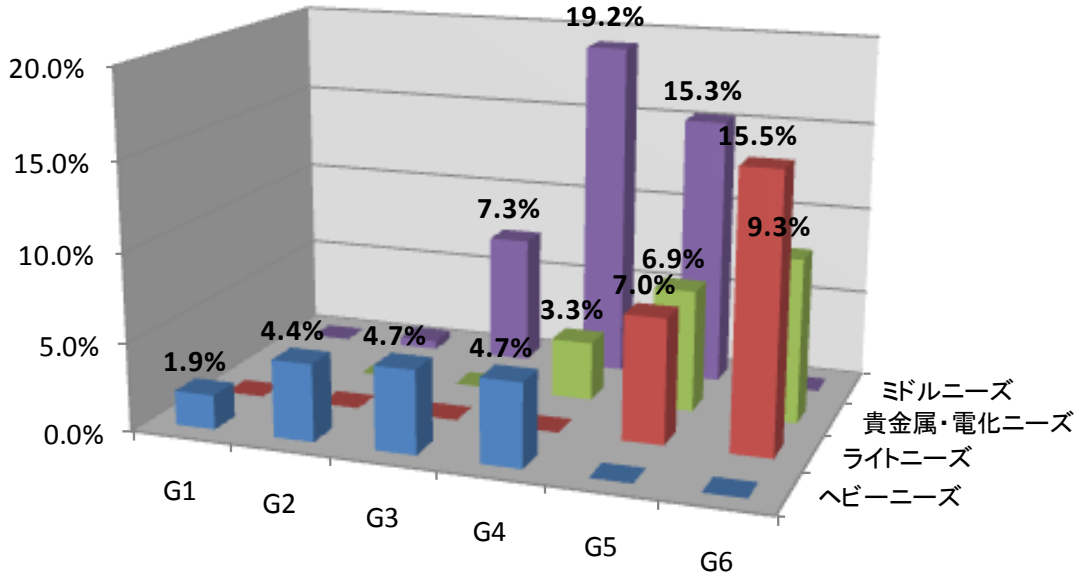


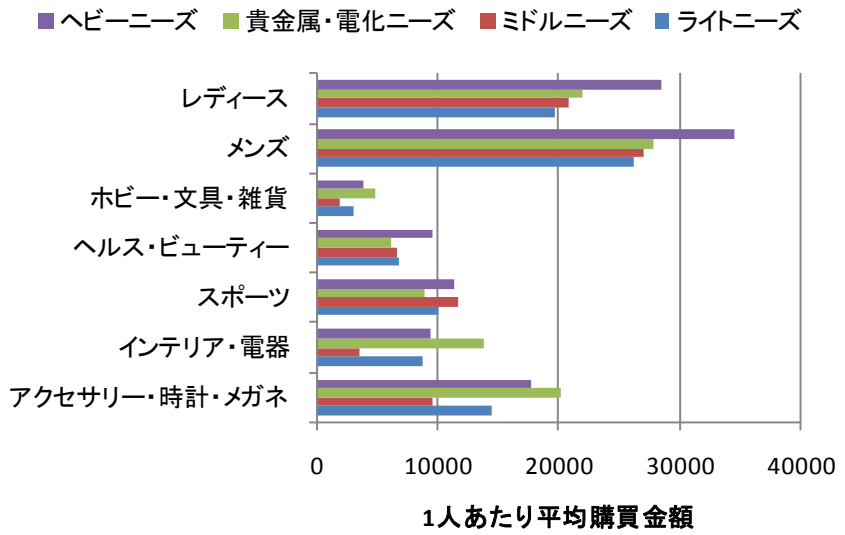


モデル1：分析結果（池袋店）

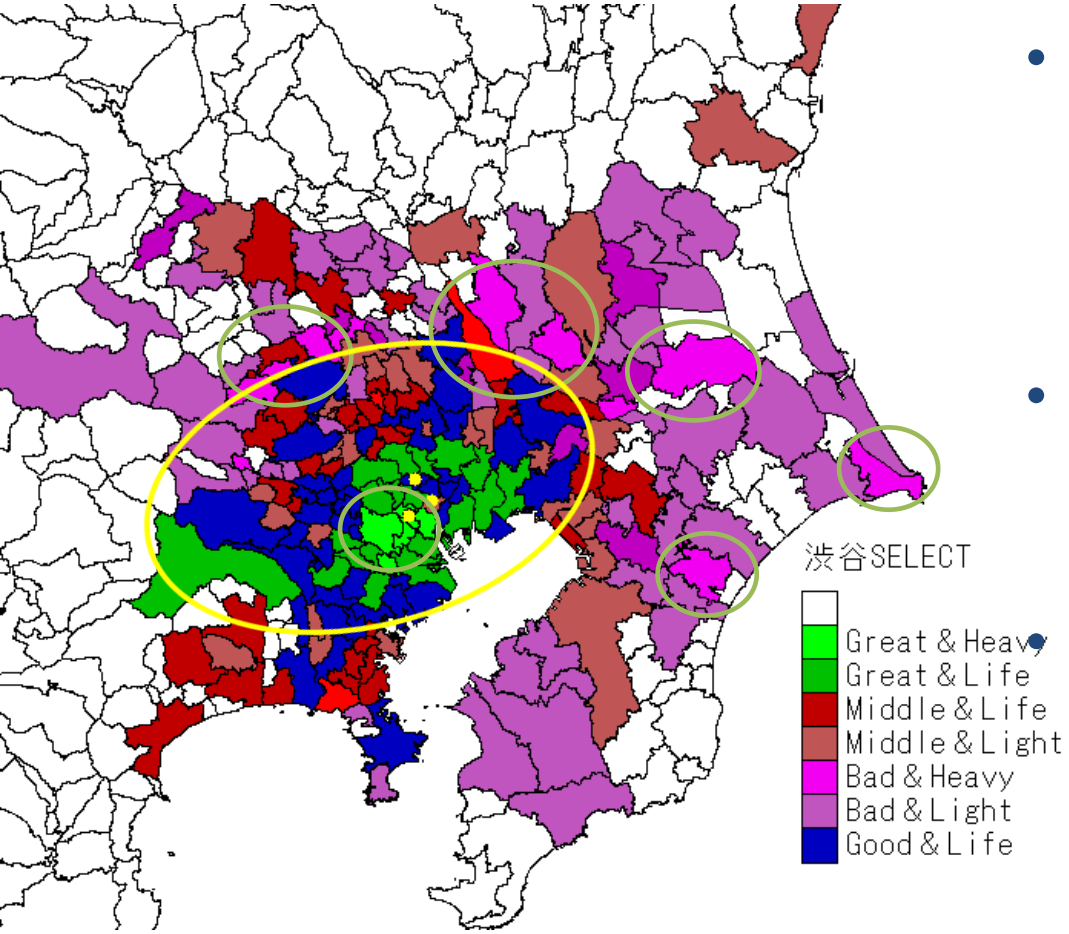
- BICの値により各セグメント基盤のセグメント数を選択
  - 顧客デモグラフィックス：6セグメント
  - 購買金額：4セグメント
  - BIC：41180.17







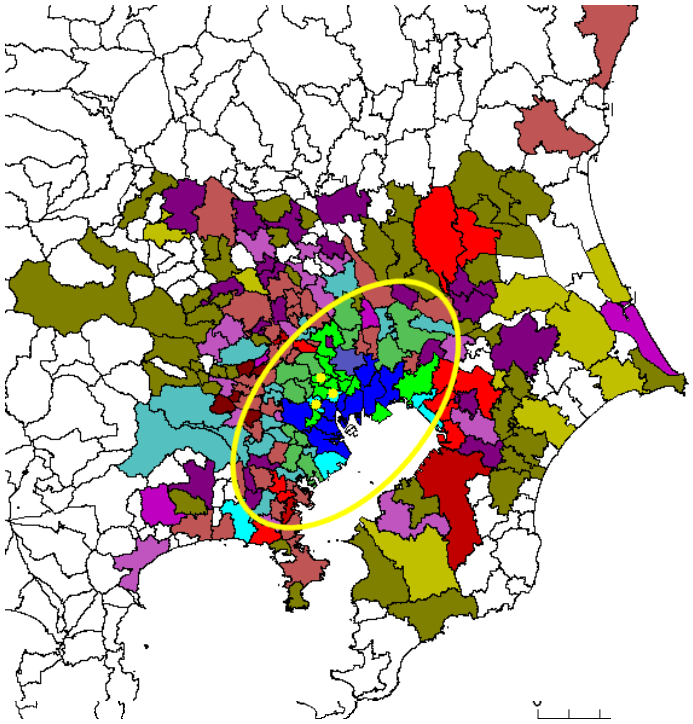
モデル1:各エリアの所属セグメント(渋谷店)



- 渋谷店は店舗から南西方向に顧客が延びている.
  - 田園都市線，東急東横線
- 顧客数は店舗からの距離に比例し，離れるほど減少していく.
- 購買金額は店舗に近い超優良エリアと店舗から遠い非優良エリアに2分される.

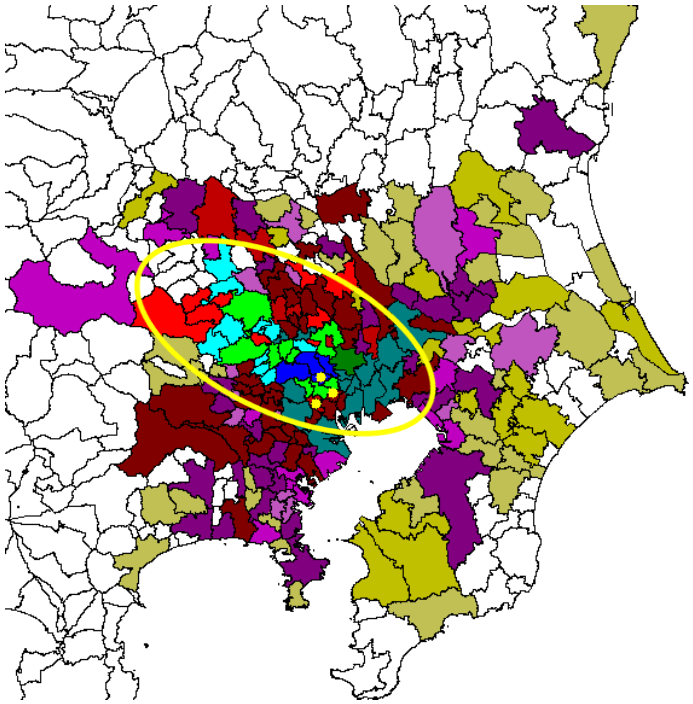
モデル1：各エリアの所属セグメント（池袋・有楽町店）

有楽町　6×4セグメント



- 東京東部を中心に東側に需要が伸びている

池袋　6×4セグメント



- 東京北西部から埼玉方向に需要が伸びている.
  - 埼京線，東武東上線，池袋線

モデル1:考察

- 顧客数は店舗からの距離に比例し，離れるほど減少していく.
- ヘビー・ニーズは店舗に近い**超優良エリア**と店舗から遠い**非優良エリア**に2分される.
  - 遠方の地域から来る顧客は，貴金属品や家電・インテリアなどの目的買いのニーズがある.

- 都内に店舗が集中しているため，優良エリアがまたがっている.
  - しかし，本モデルでは，各店舗でセグメント構造が同一でないため，店舗間の比較ができない.

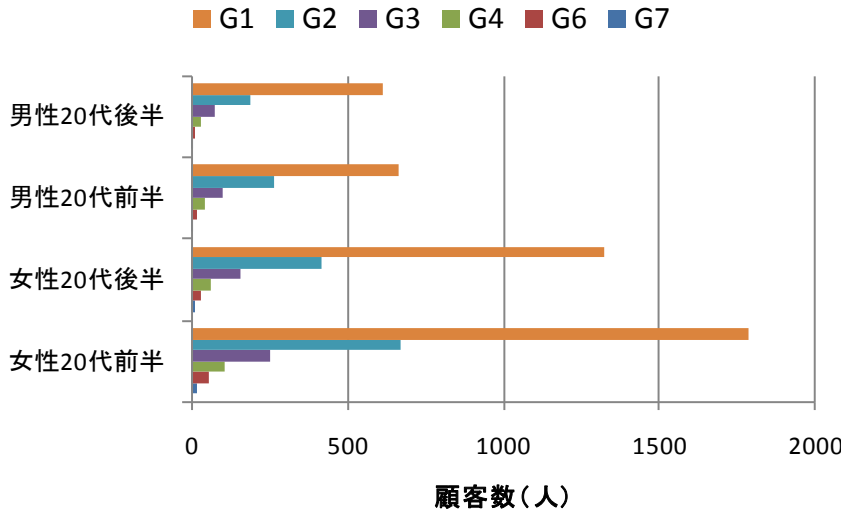
表:モデル1の店舗毎まとめ

店舗	渋谷	有楽町	池袋
セグメントサイズ	4×3	6×4	6×4
BIC	42185.21	41932.70	41180.17
実質セグメント数	9	13	12
主要顧客層	20代前半	20代後半	20代前半
購買カテゴリ特徴	メンズ レディース 貴金属類	メンズ レディース	メンズ レディース
主要エリア	首都圏南西部	首都圏東部	首都圏北東部
影響路線	東急東横線 田園都市線	–	東武東上線 埼京線 西武池袋線

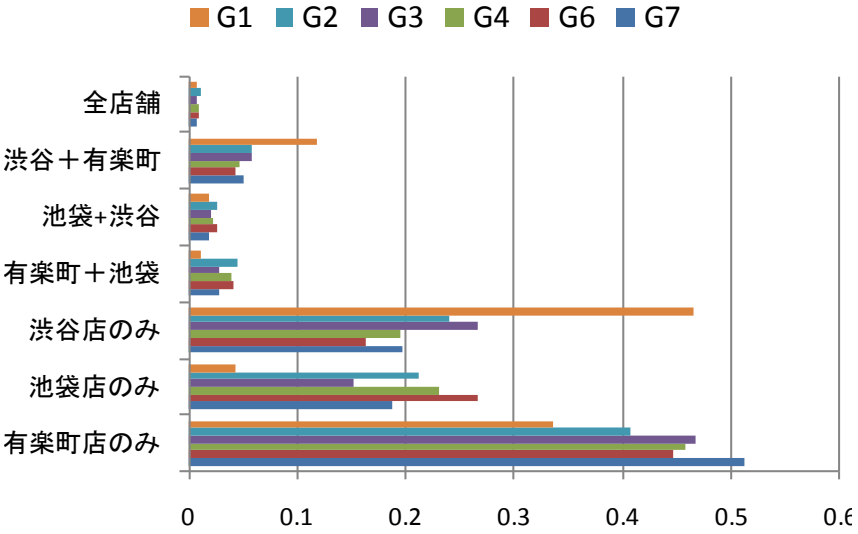
モデル2：分析結果（渋谷店）

- BICの値により各セグメント基盤のセグメント数を選択
  - 第1セグメント基盤：7セグメント
  - 第2セグメント基盤：3セグメント
  - BIC：38745.66（モデル1より良い）
- 第1セグメント基盤
  - 7つのセグメントに分類．

  - 多くのセグメントでは，有楽町店に関するθの値が高い．
    - 有楽町への店舗選択確率が高い．
    - 渋谷店の顧客は，有楽町店は併用するが，池袋店とは関係が薄い．



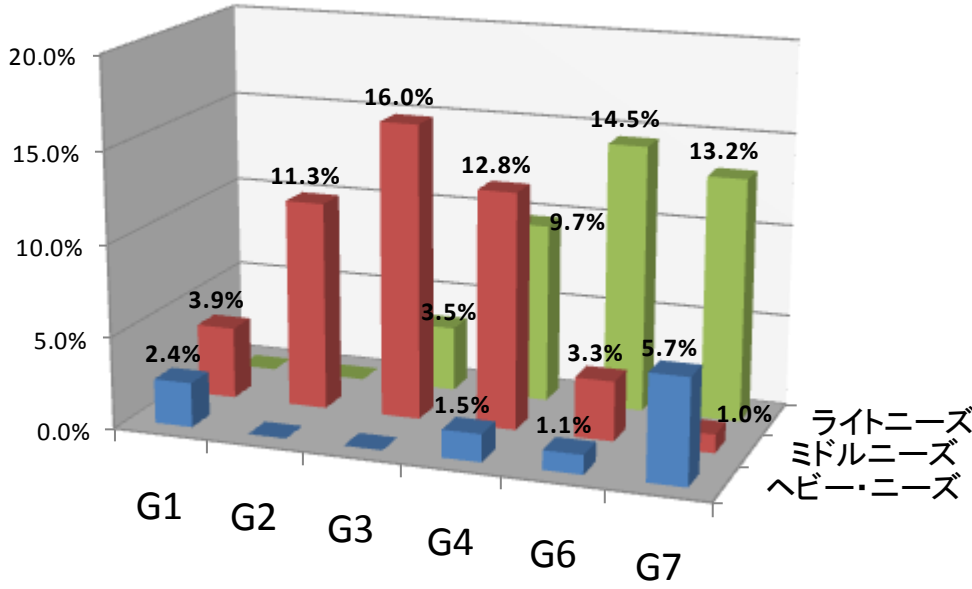
図：パラメータβ_{jd}の値



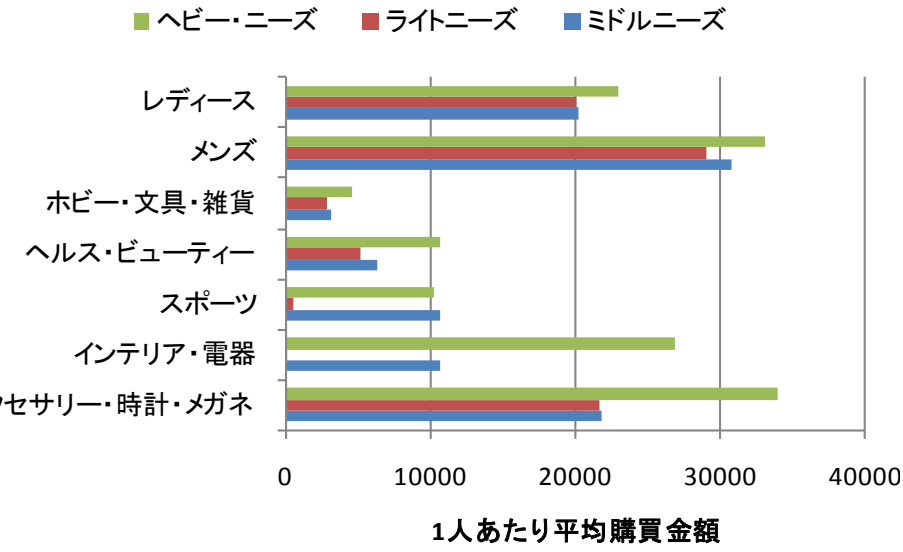
図：パラメータθ_{jc}の値

モデル2：分析結果（渋谷店）

- 第2セグメント基盤
  - モデル1同様，3つのニーズに分類.
- 所属確率
  - 実質セグメント数：17

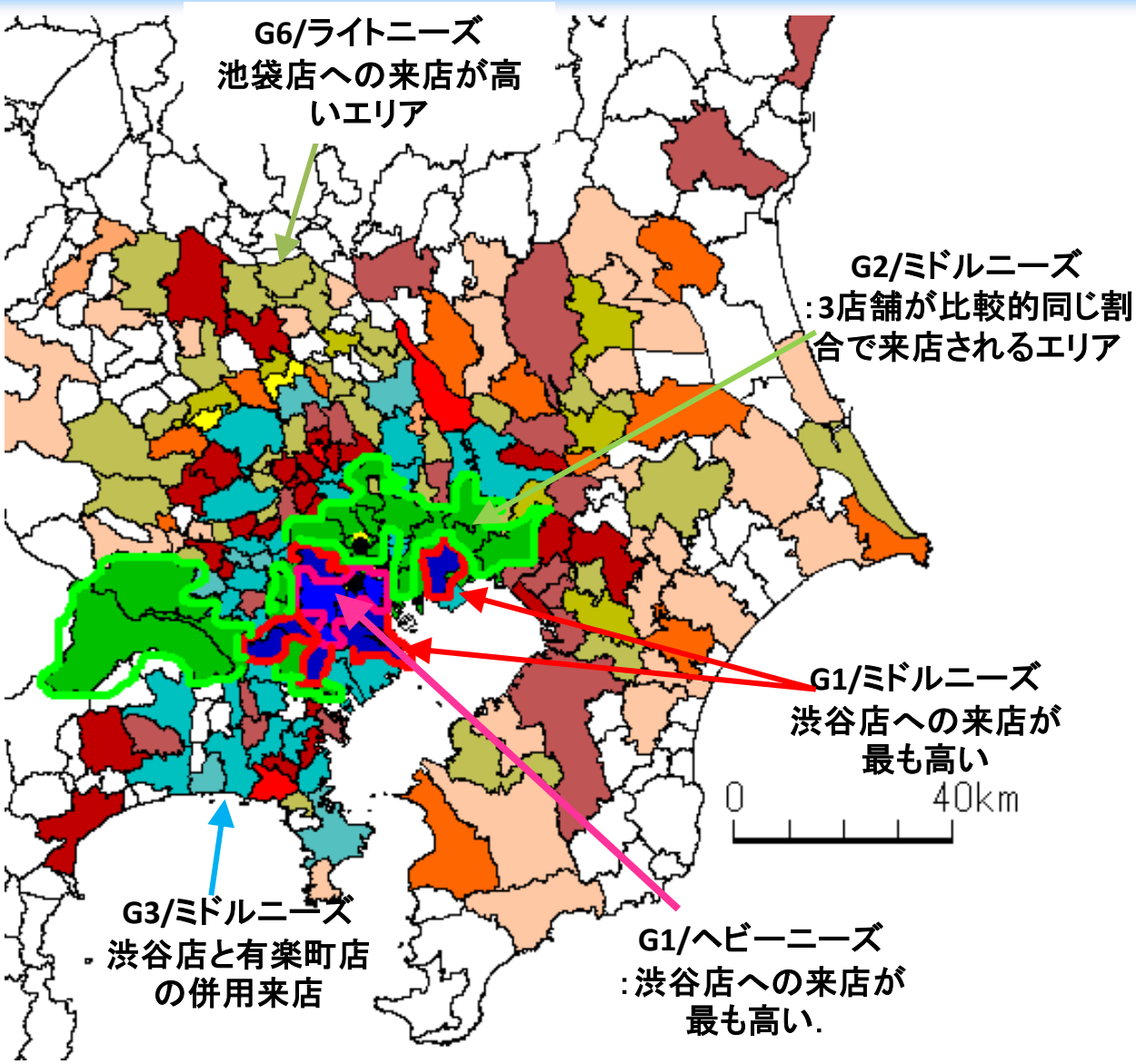


図：所属確率 φ_{jk} の値



図：パラメータ β_{ki} の値

モデル2：各エリアの所属セグメント(渋谷店)



モデル2

- ほとんどのエリアで複数店舗への来店需要がある
  - 渋谷店への顧客数が最も高いG1エリアは，ほぼ渋谷店への来店．
  - 渋谷店と有楽町店，池袋店と有楽町店の組み合わせで来店するエリアは多いが，池袋店と渋谷店の組み合わせは少ない．
  - モデル1では，優良エリアが複数店舗でまたがっているが，他店舗の影響を考慮することで，渋谷店から見た他店舗との関係を把握できた．

- 戦略策定
  - 各セグメントに合った販売促進．
  - 複数店舗へ向かわせることに，よりメリットのある販促を行うことで，競合他社が多い首都圏百貨店業界において顧客の獲得につながるのでは．
  - 今回の研究を時系列毎に分析を行い，その研究の蓄積で，需要の変化の把握を通じて，既存店舗の評価が行える．

まとめと今後の課題

- 都内百貨店を対象として，ジョイントセグメンテーションモデルを用いて，エリアセグメンテーションの形成を行い，各店舗の需要の把握を行った.
  - 商圏の設定には，顧客数の把握が重要であり，モデル1では，エリアの顧客規模をもとに分類を行った.
    - 渋谷・有楽町・池袋それぞれにおいて，需要の分布を視覚的に把握.
    - 郊外のエリアでは買い回り品の需要が見込める.
  - モデル2では，自社の他店舗もモデルに考慮して分析を行った.
    - BICはモデル1より良い値に.
    - エリアによっては，複数店舗へ顧客が分散している.

- 今後の課題
  - 対象エリアの再検証
  - 商圏モデルとの比較

参考文献

[1]里村卓也:“商圏分析のためのエリア・セグメンテーション”, オペレーションズ・リサーチ : 経営の科学 50(2), 71-76, 2005-02-01

[2]V.Ramaswamy, R.Chatterjee and H.S. Cohen:“Joint Segmentation on Distinct Interdependent Base With Categorical Data,”Journal of Marketing Reserch, Vol.33, pp.337-350,1996

[3]佐藤栄作:“商圏分析モデルの現状と課題”, オペレーションズ・リサーチ, Vol.42, No.3, pp.137-142,1997

[4]Michiko Watanabe, Kazuki Yamaguchi : *The EM Algorithm and Related Statistical Models*, Marcel Dekker, Inc

[5]Geoffrey J.McLachlan, Thriyambakam Krishnan : *The EM Algorithm and Extensions*, Wiley-Inter science

[6]小西貞則, 越智義道, 大森裕浩:「計算機統計学の方法-ブートストラップ・EMアルゴリズム・MCMC-」, 朝倉書店

[7]坂巻　英一:“個人差を考慮したジョイント・セグメンテーションモデルによる消費者セグメント構築法の提案”, Journal of the Japan Society for Management Information, Vol.11 No4, Mar.2003,pp1-15

[8]里村　卓也, 佐藤　栄作, 佐藤　忠彦:“金融商品市場へのジョイント・セグメンテーションの適応”, オペレーション・リサーチ, 2000年12月号, P631 ~636

[9] 中西正雄:「小売吸引力の理論と測定」, 千倉書房, 1983年, pp13-15

参考文献

[10]国土交通省国土計画局GISホームページ,
http://www.mlit.go.jp/kokudokeikaku/gis/index.html, 最終閲覧日 : 2010/5/18

[11] Rick L. Andrewsa, Imran S. Currim : “Recovering and profiling the true segmentation structure in markets: an empirical investigation”, Intern. J. of Research in Marketing 20 (2003) 177–192

APPENDIX

潜在クラスモデル

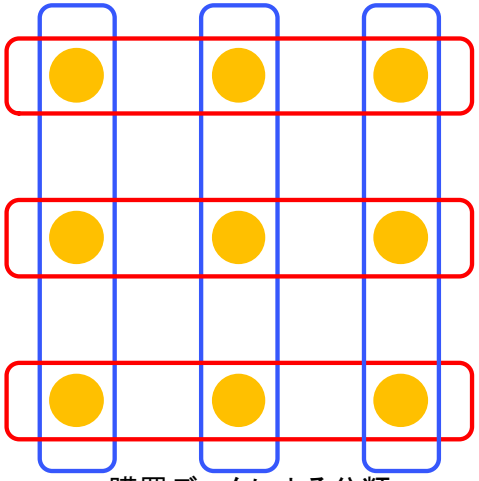
- 観測変数の背後に離散的分布をもった潜在変数を仮定し，その潜在変数の類似度に基づいてデータをクラス分けするモデル
  - ↔ 事前セグメンテーション
    顧客を事前に与えられている変数群の類似度に基づいてクラス分けする手法

- 潜在クラスモデルでは，顕在変数の間に観測される連関性は異なる応答確率に従う集団（潜在クラス）の混在が原因であるとし，各クラス内で顕在変数が互いに独立であることが仮定

- 潜在クラスモデルの主な特徴
  1. マーケットを構成する潜在クラスを決定できる
  2. 各顧客を設定した潜在クラスに分類できる
  3. 潜在クラス毎にさらなる特徴及び異質性を把握できる

ジョイント・セグメンテーションと従来のセグメント手法との比較

- クラスター分析＋クロス集計



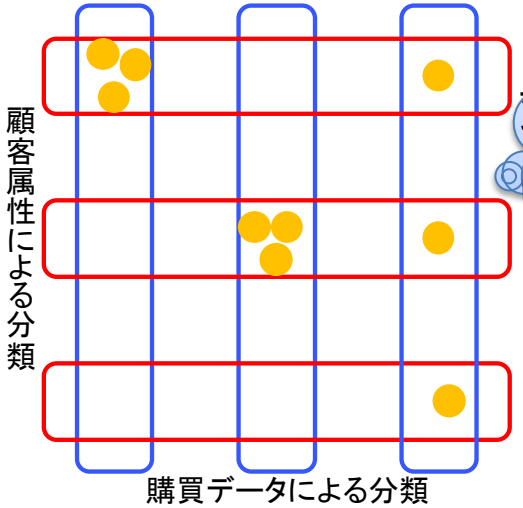
異なるセグメント軸でのクラスターは独立を仮定

- クラスターに属するか属さないか
- 主成分分析などによる情報の集約

⇒ 重要な情報損失の危険性

- クラスター数の決定には分析者の主観

- ジョイント・セグメンテーション



異なるセグメント軸での所属クラスターに相関を認める

- 所属確率の算出 ⇒ **潜在需要把握**
- 全ての情報を用いた分類

⇒ **情報の損失がない**

- **クラスター数を統計的に最も望ましい数に指定**

里村[1]のモデル

- 第1セグメント基盤：各顧客属性の顧客数
- 第2セグメント基盤：各カテゴリの平均購買金額

$$f_j(\boldsymbol{x_g}) = f_j(x_{g1},\ldots,x_{gD})$$

$$= \left(\sum_{d=1}^{D} x_{gd}\right)! \prod_{d=1}^{D}\left(\frac{\theta_{jd}^{x_{gd}}}{x_{gd}!}\right)$$

$$f_k(\boldsymbol{y_g}) = f_k(y_{g1},\ldots,y_{gI})$$

$$= \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} \exp\left(-\frac{1}{2\sigma_{ki}^2}\left(y_{gi}-\beta_{ki}\right)^2\right)$$

$f_j(\boldsymbol{x_g})$：エリアgでの第1軸（顧客属性）のセグメントjでの尤度

$f_k(\boldsymbol{y_g})$：エリアgでの第2軸（商品分類）のセグメントkでの尤度

（EMアルゴリズム）パラメータ$\phi_{jk}, \beta_{jd}, \sigma^2_{jd}, \beta_{ki}, \sigma^2_{k}$の推定

- 最尤法でも解くことが可能であるが，尤度関数が単峰ではなく複雑になっていることが多い．
- 数値的最適化によって求めた解が最大値である保証がない

EMアルゴリズム[4][5][6]の利用

•不完全なデータから形成される尤度を最大化する方法．
•E-stepとM-stepからなる操作を繰り返し行うことで欠損値を含んだモデルを推定する方法

(E-step)
観測された不完全データをいったん最尤方程式に馴染みの良い「完全データ」に疑似的に置き換える．

（M-step)
この「疑似的完全データ」を用いてパラメータの疑似最終推定値を求める．

EMアルゴリズム1

- EMアルゴリズムを利用するために潜在変数z_{gjk}を考える.
  - Z_{gjk}:エリアgがセグメントjkに所属するかの0-1変数（所属1，非所属0）

- 完全情報による対数尤度logL_C

$$L_C = \prod_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K}\left\{\phi_{jk} f_j(\boldsymbol{x_g})\cdot f_k(\boldsymbol{y_g})\right\}^{z_{gjk}} \quad \text{より，} \quad (C1)$$

$$\log L_C = \sum_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K} z_{gjk}\log\phi_{jk} + \sum_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K} z_{gjk}\log\psi_{g|jk} \quad (C2)$$

ただし，

$$\psi_{g|jk} = f_j(\boldsymbol{x_g})\cdot f_k(\boldsymbol{y_g})$$

$$= \left(\prod_{d=1}^{D}\frac{1}{\sqrt{2\pi\sigma_{jd}^2}}\exp\left(-\frac{1}{2\sigma_{jd}^2}(x_{gd}-\beta_{jd})^2\right)\right)\left(\frac{1}{\sqrt{2\pi\sigma_{ki}^2}}\exp\left(-\frac{1}{2\sigma_{ki}^2}(y_{gi}-\beta_{ki})^2\right)\right) \quad (C3)$$

EMアルゴリズム2

- E-step

  $F=(x_{gd}, y_{gi})$とパラメータ $\phi_{jk}, \beta_{jd}, \sigma^2_{jd}, \beta_{ki}, \sigma^2_{ki}$ を既知としてzについて$\log L_c$の期待値を計算.

$$E_z[\log L_C] = \sum_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K} E[z_{gjk} \mid F]\log \phi_{jk} + \sum_{g=1}^{G}\sum_{j=1}^{J}\sum_{k=1}^{K} E[z_{gjk} \mid F]\log \psi_{jk} \quad \text{(C4)}$$

  ただし, $E[z_{gjk} \mid F] = \dfrac{\phi_{jk}\psi_{g|jk}}{\sum_j \sum_k \phi_{jk}\psi_{g|jk}}$ (C5)

- M-step

  $E_z[\log L_c]$を最大とする $\phi_{jk}, \beta_{jd}, \sigma^2_{jd}, \beta_{ki}, \sigma^2_{ki}$ を求める.

  ただし,

$$0 \le \phi_{jk} \le 1, \quad \sum_{j=1}^{J}\sum_{k=1}^{K} \phi_{jk} = 1 \quad \text{(C6)}$$

h-1回とh回目での対数尤度を比較して, その差が十分に小さくなるまで繰り返す.