やさしい非集計分析

第2章

非集計分析の適用方法

2010/04/09（金）

M1　柿元淳子

内容

1. データ収集
2. 最尤推定法
3. 統計的検定
4. モデルの選択
5. 集計問題

１．データ収集

① 選択実績データ
② 各選択肢のサービス水準を表すデータ（LOSデータ）
③ 個人特性データ（年齢、性別、職業）
④ 選択を行った場や状況に関するデータで、②に含まれていないもの

1．データ収集

説明変数のタイプ

・選択肢についての変数 Z_{in}　　所要時間、費用など

・意志決定者の属性 S_n　　年齢、性別など

$$V(car) = \beta_1 + \beta_2 \text{所要時間} + \beta_4 \text{自家用車の費用} + \beta_5 \text{自家用車保有}\ (1 or 0)$$

$$V(train) = \quad \beta_2 \text{所要時間} + \beta_3 \text{公共交通運賃}$$

	β 1	β 2	β 3	β 4	β 5
自家用車	1	所要時間	0	自家用車の費用	1 or 0
鉄道	0	所要時間	公共運賃	0	0

1．データ収集

	β_1	β_2	β_3	β_4	β_5
自家用車	1	所要時間	0	自家用車の費用	1 or 0
鉄道	0	所要時間	公共運賃	0	0

例えば、自家用車の費用のパラメータと、鉄道の費用のパラメータが同じとは限らない。利用者にとって、各々の費用の重みが違うとき、別々のパラメータで考える。

しかし、選択肢間で費用の重みが変わらないときは、共通のパラメータに出来る。

２．最尤推定法

- ２つの選択肢があり、個人nが選択肢1を選んだとすると、

$$P_{1n} = P_{1n}^{\delta_{1n}} \cdot P_{2n}^{\delta_{2n}} = P_{1n}^{1} \cdot P_{2n}^{0}$$

- したがって、２つの選択肢のうち、すべての人が選んだほうの確率の積（同時確率）L*は、以下の式で表される。

$$L^* = \prod_{n=1}^{N} P_{1n}^{\delta_{1n}} \cdot P_{2n}^{\delta_{2n}}$$

2. 最尤推定法

- このとき、対数尤度関数logL*を使って解く。
- この尤度関数L*を最大にするパラメータβを推定する。すなわち最尤推定量 $\hat{\beta}$ を求める。

$$L(\beta) = \log L^*$$

$$\hat{\beta} = \arg\max_{\beta} L(\beta_1, \beta_2, \ldots, \beta_k)$$

2. 最尤推定法

- 最大値を求めるために、Lのβ_kに関する1階微分方程式 $\nabla L(\hat{\beta}) = 0$ を求めたい。

$$\nabla L(\beta) = \begin{pmatrix} \partial L / \partial \beta_1 \\ \partial L / \partial \beta_2 \\ \vdots \\ \partial L / \partial \beta_\kappa \end{pmatrix}$$

$$\nabla^2 L(\beta) = \begin{pmatrix} \partial^2 L / \partial \beta_1^2 & \partial^2 L / \partial \beta_1 \partial \beta_2 & \cdots & \partial^2 L / \partial \beta_1 \partial \beta_\kappa \\ \partial^2 L / \partial \beta_2 \partial \beta_1 & \partial^2 L / \partial \beta_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \partial^2 L / \partial \beta_\kappa \partial \beta_1 & \cdots & \cdots & \partial^2 L / \partial \beta_\kappa^2 \end{pmatrix}$$

ニュートンラプソン法

- 方程式f(x)=0を近似式 $x^{i+1} = x^i - \frac{f(x^i)}{f'(x^i)}$ により求める方法。

- すなわち、

$$\nabla L(\beta_r + \Delta\beta) = \nabla L(\beta_r) + \nabla^2 L(\beta_r) \cdot \Delta\beta = 0$$

$$\Delta\beta = -\left\{\nabla^2 L(\beta_r)\right\}^{-1} \cdot \nabla L(\beta)$$

$\beta_{r+1} = \beta_r + \Delta\beta$ として再度$\Delta\beta$を求める。

rは収束過程の繰り返し数。

$\beta_0 = 0$ から始めてこれを繰り返し、パラメータを求める。

３．統計的検定

- モデルに取り込むべき変数を取捨選択する際の判断材料

①t値検定

②尤度比ρ^2

③的中率

t値検定

- 決定したパラメータを、すべてのパラメータが0から有意に離れているかを検定する。

$$\Sigma = -\left[\nabla^2 L(\hat{\beta})\right]^{-1}$$

β_k の分散共分散行列、この行列の対角要素を $\sigma_{\kappa\kappa}^2$

$$t_k = \frac{\hat{\beta}_k}{\sqrt{\sigma_{\kappa\kappa}^2}}$$

- このときの決められた棄却限界値と比較して、有意性を判定する。
- 両側5%有意水準では、1.96以上で棄却。
- 説明変数としては通常1.5程度。

適合性

- 尤度比を用いてモデルの説明力を判断する。
- 尤度比とは、推定したパラメータによって尤度がどの程度向上したかを示す指標。
- $0<\rho^2<1$をとり、1に近づくほど、良い。

$$\rho^2 = 1 - L(\hat{\beta}) / L(0)$$

説明変数の選択

- パラメータ値の符号条件
  - Ex.)所要時間は増加すればするほど、その選択肢の効用は減ずる。したがって、所要時間のパラメータは負でなければならない。
- パラメータの有意性
  - 一定値以上のt値
- 適合性
  - 十分な尤度比ρ^2

４．モデルの選択

- 良いモデルとは･･･

①複数のモデルのパラメータが安定

②パラメータ値またはその変数の変化が被説明関数に与える感度分析結果に対する判断

③複数のパラメータに関する相対的判断

例えば、所有時間と費用のパラメータの比は時間価値（円/分）として表され、この値が経済的に妥当かどうか。

５．集計問題

- 個々の選択結果を集計する。

①数え上げ法
②積分法
③モーメント法
④分類法
⑤平均値法

実用的には、④か⑤を用いる。

分類法

- 個人をセグメント分けし、セグメントごとに説明変数の平均値を入れる。それを全体について集計。

$$S_i = \sum_{g \notin G} \frac{N_g}{N_T} \cdot S_{ig}$$

N_g セグメントgの個人数

S_{ig} そのセグメントについてのi選択肢のシェア

- セグメント　地域ごと、社会経済変数（年齢、所得など）

平均値法

- 説明変数Zの値に平均値 $\overline{Z}$ を用いる。

$$S_i = P(i / \overline{Z}, \beta)$$

S_i　i選択肢のシェア