交通行動の分析とモデリング

6.5　複数データに基づくモデル推定

6.6　離散選択モデルの応用

20100519　大村

6.5 複数データに基づくモデル推定

6.5.1　複数データソースの考え方

データには種類により利害得失がある.

・集計データと非集計データ
（大まかに把握できる or 詳しいがモデル化しにくい）

・カウントデータとアンケートデータ
（属性ないが物理量はバイアスなし or 属性あるが回答バイアスあり）

・RPデータとSPデータ
（実際の行動と属性の関係不明瞭 or 属性値変化可能だが実行動との不一致）

・異なる地域データ（地域の特徴の反映の方法）

・異なる時点データ（時点の特徴とそのつながりの捉え方）

6.5 複数データに基づくモデル推定

モデル化の方針

① あるデータソースから推定されたモデルを，別のデータによって更新する方法．

② 複数のデータソースの特徴を考慮して，モデル推定時に同時に使用する方法．

6.5.2 新たなデータによるモデルパラメータの更新

（1）一部パラメータの再推定

Ex.交通手段推定では時間価値などは等しいが，
地域特性の影響の大きい選択肢固有定数は大きく異なることが多い．

$$P_n(i) = \frac{\exp\left(\alpha_i + \mu \hat{V}_{in}\right)}{\sum_{j=1}^{J} \exp\left(\alpha_j + \mu \hat{V}_{jn}\right)} \; , \; i = 1, \ldots, J$$

新パラメータ：μ，α，$\overline{V}_{in}$は移転元での効用関数のパラメータと移転先の説明変数値からなる効用確定項

6.5 複数データに基づくモデル推定

（2）ベイズ法によるパラメータ更新方法

ある確率変数の事前分布をデータを得ることで事後分布に更新する方法.

パラメータベクトルθ，データをXとすると，

$$p(\boldsymbol{\theta} \mid \mathbf{X}) = \frac{f(\mathbf{X} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{X} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

π(θ)：θの事前分布，f(X|θ)：θが与えられたときのXの出現確率(=尤度)，p(θ|X)：Xが与えられた時のθの事後分布

あるデータソースから推定した推定量を事前情報，別のデータソースからの推定量の分布を尤度関数とみなし，両者の重み付け平均と考える.

詳しくは論文ゼミにて？すみません.

6.5 複数データに基づくモデル推定

6.5.3　RP/SPモデル

RPとSPの利害得失は補完的.

RPだけからは正確に推定できないパラメータをSPで同定すると同時に，SPのバイアスを修正.

特徴

① 2つのデータ発生過程を別々にモデル化することでSPのバイアスを需要予測から除去可能.

② 属性間のトレードオフを示すパラメータをRPとSPから同時推定することで，統計的有効性が増大する.

③ 新しいサービスの変数の係数など，RPだけからでは同定できないパラメータを推定可能.

6.5 複数データに基づくモデル推定

RPモデルとSPモデルの効用関数の多くの部分は共通．選好の方法の違いゆえに，誤差項で表される「ノイズ」の大きさは異なると考える．

RP:
$$U_{in}^{RP} = \boxed{\boldsymbol{\beta}'\mathbf{X}_{in}^{RP}} + \boldsymbol{\alpha}'\mathbf{W}_{in}^{RP} + \varepsilon_{in}^{RP} \equiv V_{in}^{RP} + \varepsilon_{in}^{RP}$$
$$(i = 1,\dots,J_n^{RP}, n = 1,\dots,N^{RP})$$

SP:
$$U_{in}^{SP} = \boxed{\boldsymbol{\beta}'\mathbf{X}_{in}^{SP}} + \boxed{\boldsymbol{\gamma}'\mathbf{Z}_{in}^{RP}} + \varepsilon_{in}^{SP} \equiv V_{in}^{SP} + \varepsilon_{in}^{SP}$$
$$(i = 1,\dots,J_n^{SP}, n = 1,\dots,N^{SP})$$

確立項の分散の関係: $Var\left(\varepsilon_{in}^{RP}\right) = \mu^2 Var\left(\varepsilon_{in}^{SP}\right) \quad ,\forall i,n$

X_{in},W_{in},Z_{in}:個人nの選択肢 i に対する説明変数ベクトル，
α，β，γ：推定パラメータ，J_n：選択肢数，N：観測数

・共通部分
・SPバイアス・SPのみの属性（新しいサービスとか）

6.5 複数データに基づくモデル推定

例. RPの誤差項にIIDガンベル分布（スケールパラメータ1）とすると左下のロジットモデル.
SPも同様にロジットモデルとすると, 右下.

$$P_n^{RP}(i) = \frac{\exp\left(V_{in}^{RP}\right)}{\sum_{j=1}^{J_n^{RP}} \exp\left(V_{jn}^{RP}\right)}, \; i = 1, \ldots, J_n^{RP} \qquad P_n^{SP}(i) = \frac{\exp\left(\mu V_{in}^{SP}\right)}{\sum_{j=1}^{J_n^{SP}} \exp\left(\mu V_{jn}^{SP}\right)}, \; i = 1, \ldots, J_n^{SP}$$

このときのパラメータ推定方法（同時推定）
それぞれの対数尤度関数

$$\ln L^{RP}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{n=1}^{N^{RP}} \sum_{i=1}^{J_n^{RP}} d_{in}^{RP} \ln P_n^{RP}(i) \qquad \ln L^{SP}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mu) = \sum_{n=1}^{N^{SP}} \sum_{i=1}^{J_n^{SP}} d_{in}^{SP} \ln P_n^{SP}(i) \qquad \ldots(1)$$

SPの対数尤度関数だけからでは, β, γとμを分離しての最大化はできないので, μ=1にする.
違う方法は, 以下の対数尤度関数を考える.

$$\ln L^{RP+SP}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mu) = \ln L^{RP}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \ln L^{SP}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mu)$$

各モデルの誤差項が独立ならば, 同時対数尤度関数となり,
これを最大化すると全てのパラメータの推定量が得られる.

6.5 複数データに基づくモデル推定

愛媛大学の倉内です.
SP/RPについては，以下にあるような段階推定は現在では
全くなされていませんので，同時推定をする必要があります.
論文発行当時は，同時推定に対応している統計パッケージが
ほとんどなかったため，段階推定の説明がなされていますが，
RやGAUSSなどであれば，簡単にできますし，作業や計算時間的
にも同時推定のほうが楽です.

同時推定といっても全く難しいことではなく，SPとRPの効用関数
およびそれぞれのモデルの尤度を同一プログラムで書いて，
最後に，全体の尤度を，RPモデルでの尤度×SPモデルでの尤度，
とするだけです.

6.5 複数データに基づくモデル推定

段階推定の推定手順

1．（1）の対数尤度関数を最大化してSPモデルのパラメータ推定値$\widehat{\mu\beta}$と$\widehat{\mu\gamma}$を得て，次式の値を計算する．$y_{in}^{RP} = \widehat{\mu\beta}'\mathbf{X}_{in}^{RP}$

2．RPモデルの効用の確定項を，$V_{in}^{RP} = \lambda \boxed{y_{in}^{RP}} + \alpha' \mathbf{W}_{in}^{RP}$ とし，これに基づくRPモデルの対数尤度関数を最大化⇒推定値λ，α

$$\hat{\mu} = \frac{1}{\hat{\lambda}},\ \hat{\beta} = \frac{\widehat{\mu\beta}}{\hat{\mu}},\ \hat{\gamma} = \frac{\widehat{\mu\gamma}}{\hat{\mu}}$$

（精度向上のために）

3．X^{sp}とZ^{sp}を$\hat{\mu}$倍してスケール変換したSPデータを作成して，RPデータとともに同時推定する．

6.5 複数データに基づくモデル推定

RP/SPモデルでは通常，同一個人からデータをとる．

↓

誤差項が独立ではなく相関がありそう

（確定項では表現できない個人の嗜好など）

このような誤差項の相関を考慮したモデルも（森川・山田，1963）

（４）モデルによる予測

実際の選択行動を表す $\hat{V}_{in} = \hat{\boldsymbol{\beta}}'\mathbf{X}_{in}^{RP} + \hat{\boldsymbol{\alpha}}'\mathbf{W}_{in}^{RP}$ を使用する．

ただし，SPモデル特有の属性があれば，それをRPモデルの効用関数に加えて予測を行う．

$$\hat{V}_{in} = \hat{\boldsymbol{\beta}}'\mathbf{X}_{in}^{RP} + \hat{\boldsymbol{\alpha}}'\mathbf{W}_{in}^{RP} + \boxed{\tilde{\boldsymbol{\gamma}}}'\boxed{\tilde{\mathbf{Z}}}_{in}^{SP}$$

↑ 部分ベクトル

6.6 離散選択モデルの応用

6.6.1 嗜好の異質性

個人ごとに係数の真値は異なるのでは？

以下は，モデル中での嗜好の異質性の表現方法.

（1）個人属性の導入

$$U_{in} = \alpha_i + \underbrace{\boldsymbol{\beta}'\mathbf{X}_{in}}_{\text{サービスレベル変数}} + \underbrace{\underbrace{\gamma_1 s_n}_{\text{性別}} + \underbrace{\gamma_2 l_n}_{\text{運転免許}}}_{\text{社会経済変数}} + \varepsilon_{in}$$

サービスレベル変数にかかる係数を個人属性で反化させることも可能.

例. 性別で単位所要時間当たりの値が異なる→ $(\beta_1 + \beta_2 s_n) t_{in}$

【短所】

個人属性の導入で嗜好の異質性を表すのは簡単であるが，どのような属性で異質性を表せるかを先験的or試行錯誤で決めなければならない.

6.6 離散選択モデルの応用

（2）マーケットセグメンテーション

嗜好がおおよそ均一だろうというサブグループに分け，
サブグループごとにパラメータ推定する．（職業とか年代とか．）

サンプル数が少ないとセグメント分割することで推定に問題が．

（3）潜在クラスモデル

個人がどのセグメントに属するかを確立モデルで表す方法．

$$P_n(i) = \sum_{s=1}^{S} P_n(i \mid s) Q_n(s) \ , \ i = 1, \ldots, J$$

$P_n(i|s)$: 個人nがセグメントsに属しているときの選択肢 i の選択確率
$Q_n(s)$: 個人nがセグメントsに属する確率（帰属確率）

セグメント総数Sを定めると，セグメントごとのパラメータや
帰属確率を推定できる．

6.6　離散選択モデルの応用

（4）個人別パラメータモデル

個人ごとにかなりの観測値が必要．パネルデータやSP調査．

（5）確率係数モデル

効用関数の係数パラメータを以下のように確率変数とする．

$$U_{in} = \overline{\boldsymbol{\beta}}'\mathbf{X}_{in} + \left(\boldsymbol{\eta}'\mathbf{X}_{in} + \varepsilon_{in}\right)$$

$$\boldsymbol{\eta} \sim \mathrm{MVN}(0, \boldsymbol{\Omega})$$

β：平均値ベクトル，Ω：共分散行列

εが多変量正規分布→多項プロビットモデル

EがIIDガンベル分布→mixed logitモデル

6.6 離散選択モデルの応用

6.2選択肢集合の考え方

（1）先験的に与えられている場合

例．自動車免許持っていなければ，自動車は除外．

（2）IIAはなりたっているが，選択肢数がきわめて大きい場合

ｉ．選択肢を集計化（目的地選択をゾーン集計にする）

真の目的地の効用値が全て同じ値の確定効用をもつとき，

$$P_n(i) = \frac{\exp\left(V_{in} + \frac{1}{\mu}\ln M_i\right)}{\sum_{j=1}^{J}\exp\left(V_{jn} + \frac{1}{\mu}\ln M_j\right)}$$

V_{in}：集計化された選択肢ｉの中の真の選択肢の確定効用値
M_i：集計化された選択肢ｉの中の要素選択肢数
確定効用値に「規模」の補正項が入っている．

ⅱ．選択肢のサンプリング

IIAの長所（選択肢集合にある全ての選択肢を考えなくても推定できる）を利用．一定数の選択肢を無作為抽出する．

6.6 離散選択モデルの応用

（3）選択肢集合生成過程のモデル化

分析者にとって選択肢集合が不確定である場合

$$P_n(i) = \sum_{C \in G} P_n(i \mid C) Q_n(C)$$

$P_n(i|C)$: 個人nの選択肢集合がCであったときの選択肢ｉ の選択確率
$Q_n(C)$: 個人n の選択肢集合がCである確率
G: 選択肢の全ての部分集合（空集合を除く）

選択肢集合形成と選択肢集合が所与の下での選択行動という2段階構成.
（選択肢の不確実性を考慮した離散選択モデル. ）

6.6.3 便益指標としてのログサム変数

（NLなどで）ツリーに含まれる全選択肢の効用の確率項がIIDガンベル分布に従い，その最頻値がいわゆるログサム変数になり，それが，前選択肢の合成効用を表している.

ガンベル分布では最頻値と期待値は0.577/μだけ異なるが，はじめから各選択肢の確率項の期待値が0になるような最頻値を決めておけば，ログサム変数は全選択肢の最大効用の期待値となる.

6.6　離散選択モデルの応用

非集計モデルにおける便益は利用可能な選択肢に関する効用の最大値の差によって計測できる.

便益　=　E（Max U_w – Max U_0）　（ただし，w:改善あり，0：改善なし）

ロジットモデルのときは，簡単に表せて，ログサム変数の差で測定される.

$$\text{便益} = \ln\left(\sum_{j \in J_n} e^{V_{jw}}\right) - \ln\left(\sum_{j \in J_n} e^{V_{j0}}\right)$$

（ただし，V_{jw}:改善ありの選択肢j の効用

V_{j0}:改善なしの選択肢j の効用）

6.6.3　便益指標としてのログサム変数

（NLなどで）ツリーに含まれる全選択肢の効用の確率項がIIDガンベル分布に従い，その最頻値がいわゆるログサム変数になり，それが，全選択肢の合成効用を表している.

ガンベル分布では最頻値と期待値は0.577/μだけ異なるが，はじめから各選択肢の確率項の期待値が0になるような最頻値を決めておけば，ログサム変数は全選択肢の最大効用の期待値となる.

6.6 離散選択モデルの応用

個人nの選択肢集合がC_nであるとき，
このログサム変数には，
選択肢の合成抵抗として望ましい2つの性質がある．

$$E\left[\max_{i\in C_n} U_{in}\right] = \frac{1}{\mu}\ln\sum_{i\in C_n}\exp(\mu V_{in})$$

①新しい選択肢が加わるとログサム変数の値は必ず大きくなる．

もし全選択肢の効用の平均値を合成効用とすると，効用値の低い選択肢が選択肢集合に加わった場合，合成効用値が下がってしまうので．

$$E\left[\max_{i\in C_n} U_{in}\right] \le E\left[\max_{i\in C'_n} U_{in}\right]$$

②選択肢の確定効用値が上がるとログサム変数の値は必ず大きくなる．

$$\frac{\partial}{\partial V_{jn}} E\left[\max_{i\in C_n} U_{in}\right] = P_n(j) > 0 \quad ,\forall j \in C_n$$

またこの式より，政策施行前後におけるログサム変数の変化は消費者余剰の変化に等しい．

$$\frac{\partial P_n(j)}{\partial V_{in}} = \frac{\partial^2 E\left[\max_{i\in C_n} U_{in}\right]}{\partial V_{in}\partial V_{jn}} = \frac{\partial P_n(i)}{\partial V_{jn}}$$

$$\Delta CS_n = \sum_{i\in C_n}\int_{\mathbf{V}_n^1}^{\mathbf{V}_n^2} P(i\mid\mathbf{V})d\mathbf{V} = \frac{1}{\mu}\ln\sum_{i\in C_n^2}\exp\left(\mu V_{in}^2\right) - \frac{1}{\mu}\ln\sum_{i\in C_n^1}\exp\left(\mu V_{in}^1\right)$$