応用計量経済学 (1): Introduction

島根哲哉

Apr. 11, 2014

What is Discrete Choice?

- ✔ 1970 年代より発達
- ✔ 質的な被説明変数を扱う回帰分析モデル
- ✔ logit model, probit model, multinomial logit model, mixed logit model, duration model, etc.
- ✔ 特に選択肢の情報による選択結果の違いに注目: conditional logit (選択肢の違い) vs. multinomial logit (意思決定主体の違い)
- ✔ 主体が選択を行う行動プロセスを明らかにすることが目的.
- ✔ 特徴 1: 効用最大化との対応 (確率効用理論)
- ✔ 特徴 2: 広い応用範囲 (産業組織論，労働経済学，環境経済学，交通計画，etc.)

Dicrete Choice Model

- ✔ 主体 n(個人，企業，意思決定者，etc.) が選択 (購入製品の決定，技術の採用，就業の継続，etc.) に直面している．
- ✔ 選択結果 y は有限個の離散的なものとする．
- ✔ 観察される要因 x と観察されない要因 ε があり，これらの要因は関数 $y = h(x, \varepsilon)$ を通じて，主体の選択と関連している．
- ✔ ε は確率的なので $h(x, \varepsilon)$ では “精確” な予測はできないが，それぞれの結果得られる確率を導くことができる．
- ✔ ε が分布 $f(\varepsilon)$ に従うとして，選択可能集合からある特定の選択肢を選び出す確率は，行動モデル

$$P(y|x) = \Pr(\varepsilon \text{ s.t } h(x, \varepsilon) = y)$$

の結果としてとらえることができる．

Dicrete Choice: Example

- ✔ スマートフォンの機種選択
  - ✘ iPhone
  - ✘ Galaxy
  - ✘ Xperia
  - ✘ 買わない
- ✔ 購入にあたって，それぞれの
  - ✘ 価格, p
  - ✘ OS, $os \in \{\text{iOS, Android}\}$
  - ✘ 通信会社, $ca \in \{\text{DoCoMo, au, Softbank}\}$
  - ✘ 機能 (処理速度，表示能力，付属ソフトウェア，etc.), f

  を検討.

Discrete Choice Example (cont.)

- ✔ 選択肢 $j \in \{\text{iPhone, Galaxy, Xperia}\}$ を選んだ時に得られると思われる効用は, $U_j = h(p_j, os_j, ca_j, f_j)$ であらわされるとする.
- ✔ 消費者は, $U_i \geq U_j, \forall j \neq i$ となる, 選択肢 i を選ぶ.
- ✔ しかし，分析者には観察されていない特性 (周囲の利用状況，販売店での取扱，デザイン，ブランドイメージ)ε_j, が購入の決め手になっているかもしれない.
- ✔ 行動モデル: $U_j = h(p_j, os_j, ca_j, f_j, \varepsilon_j)$ と定式化して推定.
- ✔ 行動モデルが正しく推定されると...
  - ✘ 需要の予測 (価格を変えたら／機能を追加したらシェアはどのように変わるか)
  - ✘ 時間の価値，機能の価値が評価

  ができる.

Probability and Integration

- ✔ 指示関数

$$I[h(x,\varepsilon)=y] = \begin{cases} 1 & \text{if } h(x,\varepsilon)=y, \\ 0 & \text{if otherwise} \end{cases}$$

を用いると,

$$P(y|x) = \Pr(I[h(x,\varepsilon)=y]=1) = \int I[h(x,\varepsilon)=y]f(\varepsilon)d\varepsilon$$

- ✔ 確率は積分である. 観察されない要因の可能な領域についての行動モデルの結果に関する指示関数の積分である.
- ✔ どのように積分を実行するか.
  - ✘ Complete Closed-From Expression
  - ✘ Complete Simulation
  - ✘ Partial Simulation, Partial Closed Form

Complete Closed-From Expression

- ✔ 特定の h と f に関しては，積分が closed form で得られる．(binary logit, logit, etc.)
- ✔ 例 (binary logit)：

$$
\begin{aligned}
U &= \beta' x + \varepsilon \ (\equiv h(x, \varepsilon)) \\
f(\varepsilon) &= \frac{\exp(-\varepsilon)}{(1+\exp(-\varepsilon))^2}, \qquad F(\varepsilon) = \frac{1}{1+\exp(-\varepsilon)} \\
P &= \int I[\beta' x + \varepsilon > 0] f(\varepsilon) d\varepsilon \\
&= \int I[\beta' x > -\varepsilon] f(\varepsilon) d\varepsilon \\
&= \int_{\varepsilon=-\beta' x}^{\infty} f(\varepsilon) d\varepsilon \\
&= 1 - F(-\beta' x) = 1 - \frac{1}{1+\exp(\beta' x)} = \frac{\exp(\beta' x)}{1+\exp(\beta' x)}
\end{aligned}
$$

Complete Simulation

- ✔ 解析的に積分が実行できない場合は，シミュレーションで数値的にこれを求めることができる．
- ✔ 分布上での積分＝平均：$\bar{t} = \int t(\varepsilon) f(\varepsilon) d\varepsilon$
- ✔ $\varepsilon^r \sim f(\varepsilon)$ をたくさん (R) 発生させたとき，$\int t(\varepsilon) f(\varepsilon) d\varepsilon \simeq \frac{1}{R} \sum_r^R t(\varepsilon^r)$
- ✔ 行動モデルに適用すれば,

1. $\varepsilon^1 \sim f(\varepsilon)$ を発生
2. $h(x, \varepsilon^1)$ を計算し，これが y であれば $I^1 = 1$, そうでなければ $I^1 = 0$.
3. $1, 2$ を R 回繰り返す．得られたものを, $I^r, r = 1, 2, ..., R$ とする．
4. I^r の平均を計算する．これがシミュレーションされた確率: $\check{P}(y|x) = \frac{1}{R} \sum_r^R I^r$.

Partial Simulation, Partial Closed Form

- ✔ ランダム項が２つ $\varepsilon = (\varepsilon_1, \varepsilon_2)$ あり，その同時分布が $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2)$ のように表わされるとする．
- ✔ 同時分布が周辺分布と条件付き分布の積, $f(\varepsilon_1, \varepsilon_2) = f(\varepsilon|\varepsilon_1) \cdot f(\varepsilon_1)$, で表わされることを用いて,

$$P(y|x) = \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon$$

$$\int_{\varepsilon_1} \left[\int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2|\varepsilon_1) d\varepsilon_2 \right] f(\varepsilon_1) d\varepsilon_1$$

のように表わせる．

- ✔ [] 内が解析的に得られ，$g(\varepsilon_1) \equiv \int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2|\varepsilon_1) d\varepsilon_2$ とする．つまり求める確率は $P(y|x) = \int_{\varepsilon_1} g(\varepsilon_1) f(\varepsilon_1) d\varepsilon_1$ となる．
- ✔ シミュレーションにより ε_1 の周辺分布について g の平均をとれば，求める確率が得られる．

Complete Closed-From Expression

- ✔ 精確な尤度を計算できる．計算に要する時間が比較的短い．
- ✔ 利用できるモデルが限られる．またそのモデルも厳しい仮定が求められる．

Complete Simulation

- ✔ どのような分布であっても，原則的に計算できる．
- ✔ 時間的な効率，精確性の向上には工夫も求められる．

Partial Simulation, Partial Closed Form

- ✔ シミュレーションだけの時よりも正確で，解析的には積分を計算できない問題についても取り組める．

この講義の進め方

授業の目標　離散選択モデルについてその実装を含めて理解し，これらのモデルを用いた論文を理解できる能力と自力でこれらのモデルを用いた分析に取り組むための能力を身につける．

評価　試験 (.6), 宿題 (.3), 授業への貢献 (.1) によって評価する．

参考資料　講義は，Kenneth E. Train(2009) “Disrcete Choice Methods with Simulation” を用いて行う．講義を理解するためには，該当箇所を通読しておくことが望ましい．このテキストは，`http://elsa.berkeley.edu/~train/` から PDF ファイルで得ることができる．

課題　5 回程度の宿題の提出を求めることを予定している．基礎的な内容の課題に，意欲のある履修生向けの課題を加える場合もあるが，基礎的課題の評価だけでも満点が取れるように取り扱う．

オフィスアワー　水曜 15:00–16:00 をオフィスアワーとします．事前に島根まで，連絡をください．また，TA の藤井さんに質問しても結構です．いずれも居室は西 8 号館 W 棟の 6F 601 号室です．

計算環境

課題を行うために，以下の計算環境を利用できるようにする必要がある．

- ✔ MATLAB(optimization toolbox, statistics toolbox を含む)
- ✔ TSUBAME 上で利用することができ，大学の研究室の PC から，Mac/Linux であれば X 端末を通じて GUI 環境を利用できる．
- ✔ Windows を利用している場合には，Cygwin を導入することで，X 端末環境を作ることができる．
- ✔ MATLAB の利用方法については講義に関連する部分についてのみ簡単に紹介するが，包括的な指導は講義の中ではできない．使い方で不明な点があれば TA まで．

授業日程

1. 4/11 イントロダクション
2. 4/18 離散選択モデルの特性 (1): Ch.2
3. 4/25 離散選択モデルの特性 (2): Ch.2
4. 5/2 数値的最大化*: Ch.8
5. 5/9 ロジットモデル (1): Ch.3
6. 5/23 ロジットモデル (2): Ch.3
7. 5/30 ネステッドロジットモデル: Ch.4
8. 6/6 乱数の抽出，プロビットモデル (1): Ch.9(9.1–2), Ch.5
9. 6/13 プロビット (2): Ch.5
10. 6/20 混合ロジット (1): Ch.6
11. 6/27 混合ロジット (2): Ch.6
12. 7/4 個人レベルのパラメータ*: Ch.11
13. 7/11 内生性 (1): Ch.13
14. 7/18 内生性 (2): Ch.13
15. 7/25 TBA

予定は進捗などに応じて変更されます.