

基礎ゼミ5

Rによる多項ロジットモデル・ ロジットモデル推定

2017/05/16

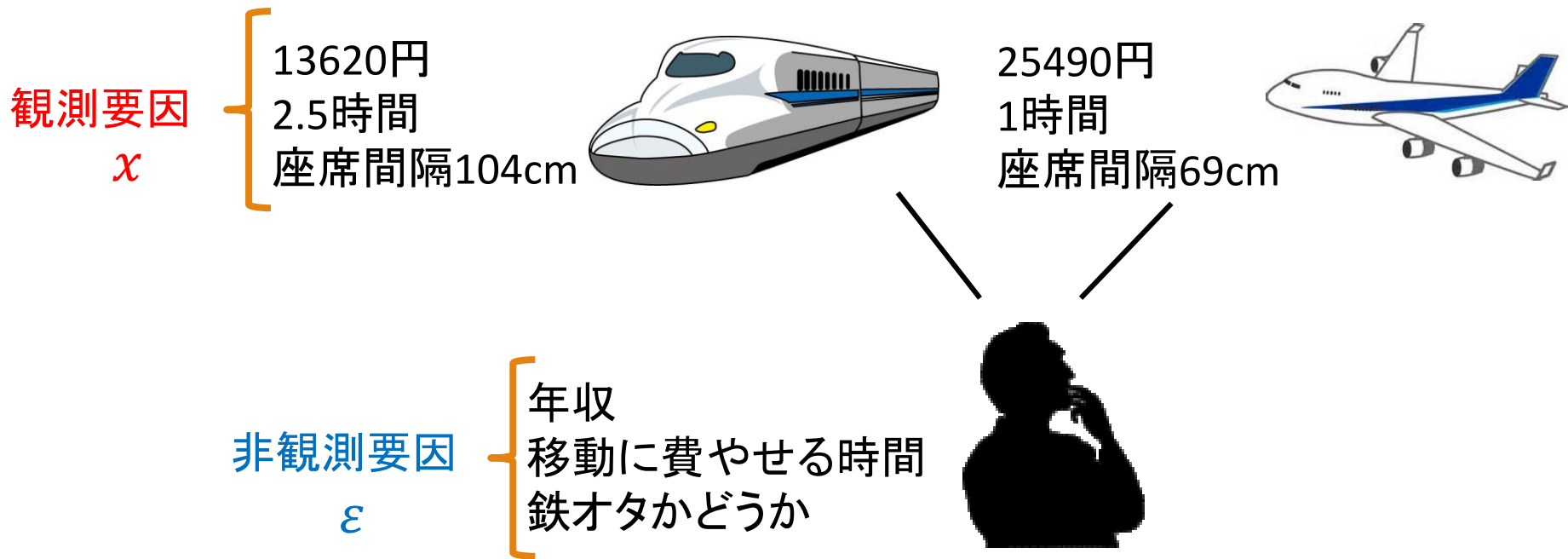
福田 研 修 士 1 年
鈴木 新

目次

- 前回の復習
- 多項ロジットモデルの推定
- ネスティッドロジットモデルの推定
- 他のモデルに関して
- 付録
- 参考文献

前回の復習

選択行動のモデル化(東京-大阪間の交通手段の例)



$$U_{(\text{選択行動})} = V(\beta x) + \varepsilon$$

← 選択行動を関数で表す(効用関数)

前回の復習

離散選択モデル

個人 n は、選択肢集合 C_n の中から、最大の効用 U_{in} を持つ選択肢を選ぶ

$$U_{in} = \underbrace{V_{in}}_{\text{確定項}} + \underbrace{\varepsilon_{in}}_{\text{ランダム項 (確率変数)}}$$

- 非観測要因(ランダム項)は観測できないため、代わりに確率変数を用いる.
- この確率変数がガンベル分布に従うモデルがロジットモデル
- 個人は選択肢集合の中から効用が最大となる選択肢を選ぶ

前回の復習

最尤推定法

・・・尤度を最大化するパラメータ値を探索する方法

例題

表が出る確率が θ であるようなコインがある。このコインを100回投げたら75回表が出た。最尤法により θ を推定せよ。

コインを100回投げたら75回表が出る確率は

$$L(\theta) = {}_{100}C_{75} \theta^{75} (1 - \theta)^{25} \quad (\text{尤度関数})$$

$L(\theta)$ を最大にする θ を考える。(計算を簡単にするため対数を取る)

$$\log L(\theta) = 75 \log \theta + 25 \log(1 - \theta) + \log {}_{100}C_{75}$$

・・・(微分とかすると)・・・

$\theta=0.75$ の時 $L(\theta)$ 最大, 尤もらしい値 ← $\theta=0.01$ でも表75回出る可能性あるけど確率低い

多項ロジットモデルの推定 (MNL)

ボーイング社における乗客の航空路選択モデル

Boeing社の路線選択調査データ, 2004年

| FEATURES | Non-Stop (Option 1) | 1 Stop (Option 2) | 1 Stop (Option 3) |
|----------------------------------|---|--|--|
| Departure time (local) | 6:00 PM | 4:30 PM | 6:00 PM |
| Arrival time (local) | 8:14 PM | 8:44 PM | 9:44 PM |
| Total time in air | 4 hr 14 min | 4 hr 44 min | 4 hr 44 min |
| Total trip time | 4 hr 14 min | 6 hr 14 min | 5 hr 44 min |
| Legroom <input type="checkbox"/> | typical legroom | 2-in more of legroom | 4-in more of legroom |
| Airline [Airplane] | Depart Chicago Continental Airlines [B737] to San Diego | Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego | Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego |
| Fare | \$565 | \$485 | \$620 |
| 1. Which is MOST attractive? | <input checked="" type="radio"/> Option 1 | <input type="radio"/> Option 2 | <input type="radio"/> Option 3 |

Option1 (直行便),
Option2 (他空港立ち寄り, 同機体),
Option3 (他空港立ち寄り, 乗り換え)

最も魅力的な路線を
1つ選んでもらう

多項ロジットモデルの推定 (MNL)

●データの項目

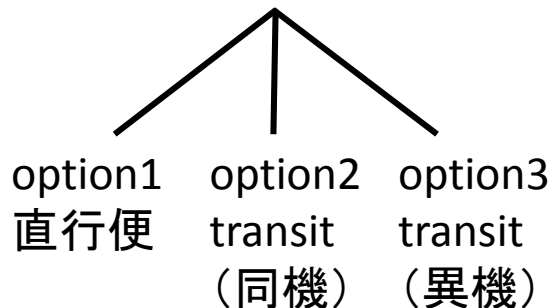
| 項目 | 内容 | 項目 | 内容 |
|---------------------|-------------|---------------------------|--------------------|
| Subj_Age | 年齢 | SP1_MostAttractive | 選択したルート |
| Subj_Occupation | 職業 | <u>OptX_Fare</u> | <u>料金(\$)</u> |
| Subj_Income | 収入(\$) | <u>OptX_TotalTriptime</u> | <u>移動時間(hours)</u> |
| Trip_Purpose | 移動目的 | Trip_PartySize | 同行者数 |
| <u>OptX_Legroom</u> | <u>座席間隔</u> | Trip_Miles | 移動距離(miles) |

今回, Option1, Option2, Option3の選択において, それぞれのルートの効用関数を考え, 各項目のパラメータを推定する.

多項ロジットモデルの推定(MNL)

今回、「移動時間」、「料金」、「座席間隔」の項目を使用すると、Option1, Option2, Option3の各効用関数 V_{op} は次のようになる。

$$\begin{aligned} V_{op1} &= ACS_{op1} + \beta_{time} \cdot x_{op1_time} + \beta_{fare} \cdot x_{op1_fare} + \beta_{leg} \cdot x_{op1_leg} \\ V_{op2} &= ACS_{op2} + \beta_{time} \cdot x_{op2_time} + \beta_{fare} \cdot x_{op2_fare} + \beta_{leg} \cdot x_{op2_leg} \\ V_{op3} &= \beta_{time} \cdot x_{op3_time} + \beta_{fare} \cdot x_{op3_fare} + \beta_{leg} \cdot x_{op3_leg} \end{aligned}$$



ACS : 選択肢固有定数(定数項)

β_{time} : 所要時間に対するパラメータ

β_{fare} : 料金に対するパラメータ

β_{leg} : 座席間隔に対するパラメータ(=快適さの指標)

データの読み込み

```
1 ### Multinomial Logit model estimation
2 setwd("C:/Users/a.suzuki/Dropbox (fukudalab-tokyotech)/junryu")
3 #データファイルの読み込み
4 Data <- read.csv("boeing.csv",header=T)
5 #データの行数(サンプル数)を数える
6 nos <- nrow(Data)#nosにサンプル数を代入
7 #パラメータの初期値設定, 0ベクトル
8 b0 <- numeric(5)#パラメータ5個あるから5
9
```

Dataの行数を調べる

要素が0のベクトルを作成する
c(0,0,0,0,0)

尤度関数の定義，初期値の設定

```
17 fr <- function(x) {  
18   #定数項  
19   b1 <- x[1]  
20   b2 <- x[2]  
21   #所要時間  
22   t1 <- x[3]  
23   #料金  
24   f1 <- x[4]  
25   #座席間隔  
26   L1 <- x[5]  
27   #対数尤度のための変数を宣言  
28   LL = 0  
29  
30 }
```

Function関数で尤度関数を定義する
※function関数の例

`a <- function(x){x+5}`
a(2)でまわすと7が出る

変数の定義

$b1 = x[1] = ACS_{op1}$
 $b2 = x[2] = ACS_{op2}$
 $t1 = x[3] = x_{time}$

$f1 = x[4] = x_{fare}$
 $L1 = x[5] = x_{leg}$

尤度をLLと定義し，初期値を0に

効用関数の定義

$$\begin{aligned}V_{op1} &= ACS_{op1} + \beta_{op1_time} \cdot x_{time} + \beta_{op1_fare} \cdot x_{fare} + \beta_{op1_leg} \cdot x_{leg} \\V_{op2} &= ACS_{op2} + \beta_{op2_time} \cdot x_{time} + \beta_{op2_fare} \cdot x_{fare} + \beta_{op2_leg} \cdot x_{leg} \\V_{op3} &= \beta_{op3_time} \cdot x_{time} + \beta_{op3_fare} \cdot x_{fare} + \beta_{op3_leg} \cdot x_{leg}\end{aligned}$$

移動時間

料金

```
option1 <- exp(t1*Data$Opt1_TotalTriptime + f1*Data$Opt1_Fare/100
option2 <- exp(t1*Data$Opt2_TotalTriptime + f1*Data$Opt2_Fare/100
option3 <- exp(t1*Data$Opt3_TotalTriptime + f1*Data$Opt3_Fare/100
```

座席間隔

定数項

```
+ L1*Data$Opt1_Legroom + b1*matrix(1,nrow =nos,ncol=1))
+ L1*Data$Opt2_Legroom + b2*matrix(1,nrow =nos,ncol=1))
+ L1*Data$Opt3_Legroom)|
```

サンプル数×1のマトリックス

料金については、説明変数のおよそのケタを揃えて推定計算を安定させるため、100で割り100\$を単位として変数をとる。

選択確率の導出および選択結果の変換

各Optionの選択確率, 例えばOption1の場合は

$$P_{op1} = \frac{e^{V_{op1}}}{e^{V_{op1}} + e^{V_{op2}} + e^{V_{op3}}}$$

```
38 #分母となる各々のEXP(v)の和を作る
39 deno <- (option1 + option2 + option3)
40 #それぞれ計算する
41 Poption1 <- (option1 / deno)
42 Poption2 <- (option2 / deno)
43 Poption3 <- (option3 / deno)
..
```

選択結果の数値変換

```
45 #選択結果 (True or False だが1,0と同じ)
46 Coption1 <- Data$SP1_MostAttractive == "1"
47 Coption2 <- Data$SP1_MostAttractive == "2"
48 Coption3 <- Data$SP1_MostAttractive == "3"
```

データのSP1_MostAttractive(選択結果)の列が「Option3」の場合は1を, そうでない行に関しては0を出力する.

対数尤度関数の設定

ルート3を選択した場合, ルート3の選択確率が算出される
ex) Option1を選択したデータの場合

$$\rightarrow 1 * \log\left(\frac{e^{V_{op1}}}{e^{V_{op1}} + e^{V_{op2}} + e^{V_{op3}}}\right)$$

```
49  
50 #対数尤度関数(ここを後に最大化する)  
51 LL <- colSums(Croute1*log(Proute1) + Croute2*log(Proute2) +  
52               Croute3*log(Proute3))  
53 }
```

その和をとることで尤度関数を算出

尤度関数の定義は終了, この関数を最大化する.

対数尤度関数の最大化

```
55 ▾ ##### 対数尤度関数の最大化 #####
56
57 ##尤度を最大化する係数の推定, パラメーター値の最適化
58 res <- optim(b0,fr, method = "Nelder-Mead", hessian = TRUE, control=list(fnscale=-1))
```

Optim・・・最適化関数

対数尤度関数(fr)を最適化(尤度を最大化)する

パラメーターを求める

```
> print(res)
$par
[1] 1.2344185 0.1067678 -0.4150240 -1.9071265 0.1967743

$value
[1] -1082.261

$counts
function gradient
      478         NA

$convergence
[1] 0

$message
NULL

$hessian
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -272.57839 140.73952 462.0639 -83.112883 82.858210
[2,] 140.73952 -176.91895 -238.3211 39.468706 -35.740418
[3,] 462.06387 -238.32107 -888.2352 151.741911 -134.564903
[4,] -83.11288 39.46871 151.7419 -125.718954 8.417851
[5,] 82.85821 -35.74042 -134.5649 8.417851 -796.398185
```

\$par

パラメーターの最適解の値

\$value

最大尤度

\$message

エラーのありなし, NULLなら無し

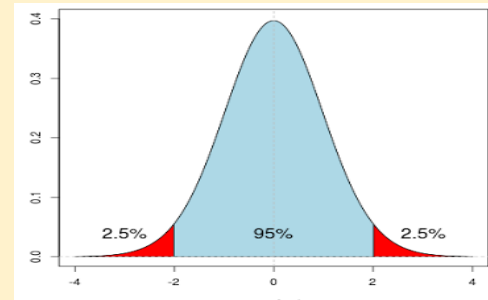
\$hessian

ヘッセ行列(t値の計算に必要)

t値の算出

```
61  
62 ## パラメータ推定値，ヘッセ行列  
63 b <- res$par  
64 hhh <- res$hessian  
65  
66 ## t値の計算  
67 tval <- b/sqrt(-diag(solve(hhh)))  
68
```

$$t\text{値} = \left| \frac{\beta}{\sigma} \right|$$



solve・・・逆行列を算出する。

diag・・・行列の作成

sqrt・・・平方根の計算

ヘッセ行列を分散共分散行列に

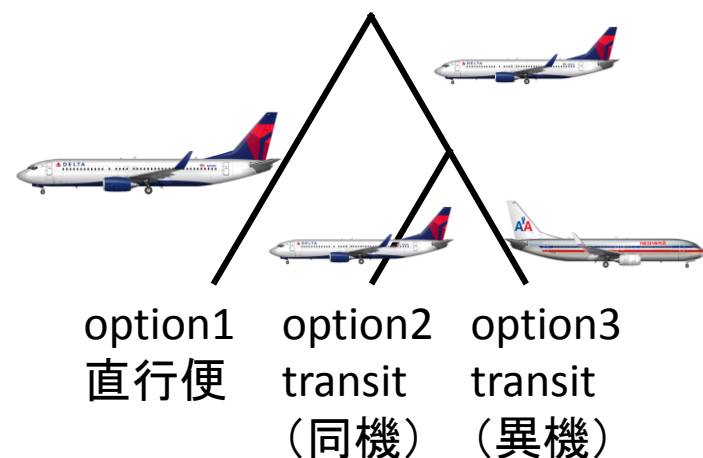
標準偏差の算出

最尤推定量を標準偏差で割ることでt値を算出

t値が1.96以上なら有意水準5%を満たす

ネステッドロジットモデルの推定(NL)

- MNLの限界 (IIA特性)
→ NL (ネステッドロジットモデル)
- 今回は直行便かTransit便かで分ける.
- プログラム
 - 大体MNLと同じ
 - パラメーターの初期設定や尤度関数など一部変更あり



ネステッドロジットモデルの推定(NL)

```
88 nos <- nrow(Data)#nosにサンプル数を代入
89 #パラメータの初期値設定
90 b0 <- c(0,0,0,0,0,1)#パラメータ6個あるから6
91
```

```
93 fr <- function(x) {
94   #定数項
95   b1 <- x[1]
96   b2 <- x[2]
97   #所要時間
98   t1 <- x[3]
99   #料金
100  f1 <- x[4]
101  #座席間隔
102  L1 <- x[5]
103  #スケールパラメーター
104  pa <- x[6]
105  #対数尤度のための変数を宣言
```

スケールパラメーター分
1個増やす(初期値は1に)

関数にもスケールパラメーターを定義

ネステッドロジットモデルの推定(NL)

```

111 option3 <- exp(t1*Data$Opt3_TotalTriptime + f1*Dat
112
113 #ネストを作る, 直行便かそうでないかの選択確率(ツリ
114 nume.direct <- exp(pa*log(option1))
115 nume.transit <- exp(pa*log(option2+option3))
116
117 deno <- nume.direct + nume.transit
118 P.direct <- nume.direct / deno
119 P.transit <- nume.transit / deno
120

```

ツリー上部の選択確率の作成
(直行便かどうか, ログサム変数)

$$\dot{V}_i = \frac{1}{\beta_i} \ln \left(\sum_{i \in C} \exp(\beta_i V_i) \right)$$

$$P(m) = \frac{\exp(\mu_m \dot{V}_l)}{\sum_l \exp(\mu_m \dot{V}_l)}$$

i:op1, op2, op3
m:direct, transit

```

121
122 #次に, transitのなかでの選択, 条件付き確率を考える.(ツリー下部
123 #P(LV2|LV1)
124 ## 直行便のなかでの選択(1個しかないからそのまま)
125 deno.direct <- option1
126 P.option1.direct <- option1 / ((deno.direct!=0)*deno.direct + (deno.direct==0))
127
128 ## transit便のなかでの選択
129 deno.transit <- option2+option3
130 P.option2.transit <- option2 / ((deno.transit!=0)*deno.transit + (deno.transit==0))
131 P.option3.transit <- option3 / ((deno.transit!=0)*deno.transit + (deno.transit==0))
132

```

ツリー下部においても同様, 条件付き確率を算出
(transitの場合同機か異機か)

$$P(i|m) = \frac{\exp(\mu_m V_i)}{\sum_j \exp(\mu_m V_j)}$$

ネステッドロジットモデルの推定(NL)

```
133  
134   ### 最後に、それぞれの交通手段の同時確率を計算するP(  
135   P.option1 <- P.option1.direct * P.direct  
136   P.option2  <- P.option2.transit  * P.transit  
137   P.option3  <- P.option3.transit * P.transit
```

同時確率を算出

$$P(i, m) = P(i|m)P(m) = \frac{\exp(\mu_m V_i)}{\sum_j \exp(\mu_m V_j)} \frac{\exp(\mu_m \dot{V}_l)}{\sum_l \exp(\mu_m \dot{V}_l)}$$

```
150  
151 res <- optim(b0,fr, method = "BFGS", hessian = TRUE, control=list(fnscale=-1))  
152  
153
```

MethodをBFGSに

他のモデルに関して (CNL, MXL)

下記のサイトにコードが載っているので、参考にしてください。

<http://bin.t.u-tokyo.ac.jp/kaken>

付録：データの補足説明

| Variable | Description |
|--------------------|--|
| Subj_ID | Unique identifier for each respondent. |
| Subj_Male | 1 if male, 0 otherwise |
| Subj_Age | Age in years |
| Subj_Occupation | Occupation (01 = Executive and Managerial, 02 = Professional, 03 = Technicians and related support, 04 = Sales, 05 = Administrative support, 06 = Services, 07 = Precision production, craft, repair, 08 = Machine operators, assemblers, inspectors, 09 = Transportation and material moving, 10 = Handlers, cleaners, helpers, 11 = Farming, forestry, and fishing, 12 = Armed forces) |
| Subj_Income | Annual income in \$ |
| Subj_IncomeMissing | Income is missing |
| Subj_Education | Education (01 = Less than High School Diploma, 02 = High School Graduate, 03 = Some college, No Degree, 04 = Associate Degree - Occupational, 05 = Associate Degree - Academic, 06 = Bachelors Degree, 07 = Masters Degree, 08 = Professional Degree, 09 = Doctorate Degree) |

| Variable | Description |
|-------------------------|--|
| Trip_Purpose | Trip purpose (1=business, 2=leisure, 3=attending conference/seminar/training, 4=both business and leisure) |
| Trip_TravelerPays | 1 if the traveler is paying for the trip, 0 if someone else is paying |
| Trip_IdealDepartureTime | Respondents ideal departure time (hours after midnight) |
| Trip_PartySize | Number of persons traveling |
| Trip_OrigMinGMT | Origin city time zone (minutes from GMT) |
| Trip_DestMinGMT | Destination city time zone (minutes from GMT) |
| Trip_BaseFlightTime | Flight time for shortest non-stop itinerary in minutes |
| Trip_Miles | Length of itinerary in miles |
| Trip_Direction | Direction of itinerary (1=East to West, 2=West to East, 3=North-South) |

| Variable | Description |
|----------------------|--|
| OptX_DepTimeHrs | Option X: Departure time, local (hours after midnight) |
| OptX_ArrTimeHrs | Option X: Arrival time, local (hours after midnight) |
| OptX_TotalTimeInAir | Option X: Total time in air (hours) |
| OptX_TotalTriptime | Option X: Total trip time (hours) |
| OptX_Legroom | Option X: Legroom (1 = 2 inches less than typical, 2 = typical, 3 = 2 inches more than typical, 4 = 4 inches more than typical) |
| OptX_AirlineA | Option X: Airline for first leg (only known to arbitrary airline number for proprietary reasons) |
| OptX_AirlineB | Option X: Airline for second leg (if there exists a second leg) (only known to arbitrary airline number for proprietary reasons) |
| OptX_AirplaneA | Option X: Airplane for first leg (only known to arbitrary airplane number for proprietary reasons) |
| OptX_AirplaneB | Option X: Airplane for second leg (if there exists a second leg) (only known to arbitrary airplane number for proprietary reasons) |
| OptX_Fare | Option X: Fare (\$) |
| OptX_SchedDelayEarly | Option X: Schedule delay (hours) - early departure (calculated from OptX_DepTimeHrs and Trip_IdealDepartureTime) |
| OptX_SchedDelayLate | Option X: Schedule delay (hours) - late departure (calculated from OptX_DepTimeHrs and Trip_IdealDepartureTime) |