2016/05/18

福田研･朝倉研合同ゼミ 交通行動分析ゼミ 第2回②

第３部 モデリング

･動的モデル
･生存時間モデル

朝倉研B4 山下優希

目次

- 動的モデル

  交通行動の動的特性

  確率過程としての交通行動

  離散時間パネルデータの解析①②

  パネルデータの有効性

- 生存時間モデル

  基礎概念

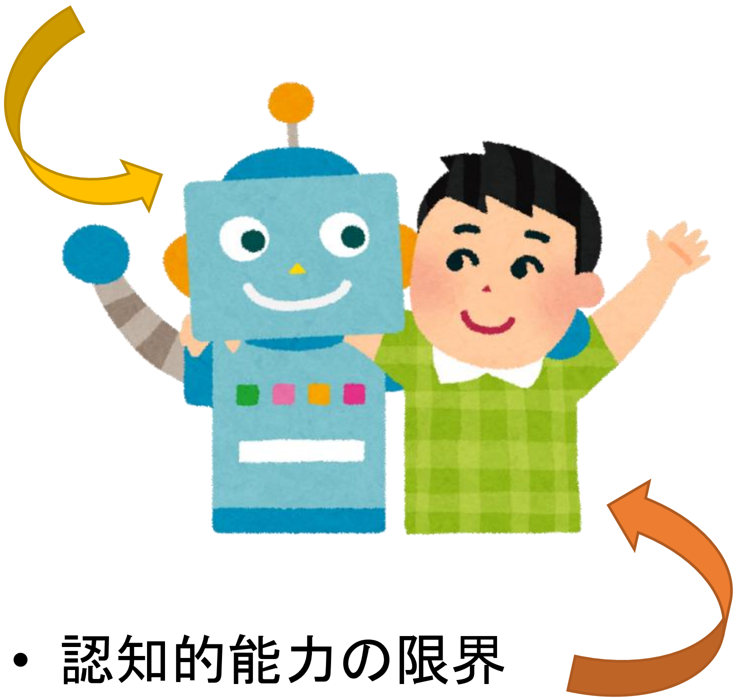
  生存時間の解析方法

  基本モデルの拡張

  適用事例

交通行動の動的特性

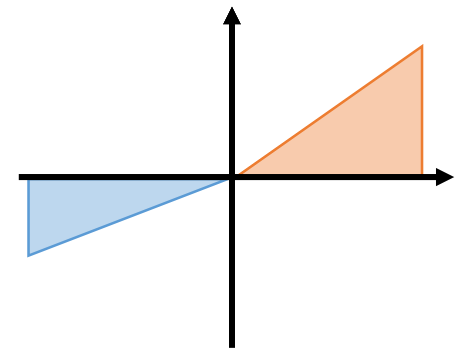
- 行動は均衡状態→静的な分析



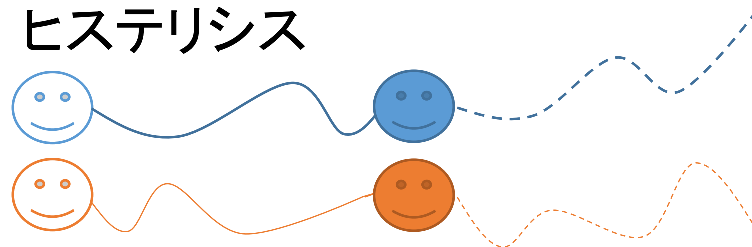
- 認知的能力の限界
- 不完全情報
- 行動適応に伴う時間ずれ

→行動変化の過程に着目

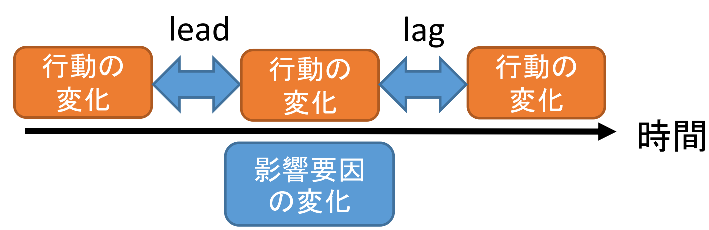
- 非対称性



- ヒステリシス



- 時間ずれ



確率過程としての交通行動

時間とともに変化する確率変数

交通行動が確率過程として記述されるとし、数学的な定式化をし、頻繁に用いられる確率過程モデルについて述べる

- 再生過程
- マルコフ再生過程
- マルコフ過程
- マルコフ連鎖

確率過程を $\mathrm{T} = \{T_n; n \in \mathrm{N}_+\}$

滞在時間が同一の分布を持ちかつ互いに独立

滞在時間の累積分布　$\mathrm{F_T}(t)$

$$P_r[T_{n+1} - T_n \leq t | T_0, \dots, T_n] = P_r[T_{n+1} - T_n \leq t\,] = \mathrm{F_T}(t)$$
$$t > 0, \forall n \in \mathrm{N}_+$$

状態空間 E　非負の整数の集合 N_+

$\mathrm{R}_+ = [0,+\infty]$

n番目の遷移後の状態 $\mathrm{X}_n, n \in \mathrm{N}_+$

n番目の遷移が生じる時刻 T_n

$\mathrm{X}_n \in E, T_n \in \mathrm{R}_+, 0 = T_0 \leq T_1 \leq T_2 \cdots$

この確率過程を$(\mathrm{X}, \mathrm{T}) = \{\mathrm{X}_n, T_n; n \in \mathrm{N}_+\}$

例）交通事故の発生、

電話回線での通話の開始

確率過程としての交通行動

交通行動が確率過程として記述されるとし、数学的な定式化をし、頻繁に用いられる確率過程モデルについて述べる

複数状態

- 再生過程
- **マルコフ再生過程**
- マルコフ過程
- マルコフ連鎖

状態空間 E 非負の整数の集合 N_+

$R_+ = [0,+\infty]$

n番目の遷移後の状態 $X_n, n \in N_+$

n番目の遷移が生じる時刻 T_n

$X_n \in E, T_n \in R_+, 0=T_0 \le T_1 \le T_2 \cdots$

この確率過程を$(X, T) = \{X_n, T_n; n \in N_+\}$

確率過程を(Z, T)

$\forall n \in N_+, j \in E, t \in R$について

$$P_r[X_{n+1} = j, T_{n+1} - T_n \le t | X_0, \dots, X_n; T_0, \dots, T_n] = P_r[X_{n+1} = j, T_{n+1} - T_n \le t | X_n]$$

さらに$i, j \in E, t \in R$について

$$P_r[X_{n+1} = j, T_{n+1} - T_n \le t | X_n = i] = Q(i,j,t)$$

が成立すると仮定する（時間的一様性）

状態空間Eのセミ・マルコフ核

確率群$Q = \{Q(i,j,t) \ : i, j \in E, \ t \in R\}$

状態iから状態jへの遷移確率

$$P(i,j) = \lim_{t\to\infty} Q(i,j,t)$$

確率過程としての交通行動

交通行動が確率過程として記述されるとし、数学的な定式化をし、頻繁に用いられる確率過程モデルについて述べる

- 再生過程
- マルコフ再生過程
- **マルコフ過程**　条件付き過去独立
- マルコフ連鎖

時間的一様なマルコフ過程のQは

$Q(i,j,t) = P(i,j)\left(1 - e^{-\lambda_i t}\right), t \geq 0, \forall \mathrm{i}, \mathrm{j} \in E$

状態i, $(\forall \mathrm{i} \in E)$での滞在時間は

パラメーターλ$_\mathrm{i}$の負の指数分布を持ち、

状態間の遷移確率$P(i,j)$は

状態iでの滞在時間から独立

ここでは$P(i,j) = 0, \forall \mathrm{i} \in E$

を仮定する

状態空間E　非負の整数の集合 N_+

$\mathrm{R}_+ = [0,+\infty]$

n番目の遷移後の状態 $\mathrm{X}_n, n \in \mathrm{N}_+$

n番目の遷移が生じる時刻 T_n

$\mathrm{X}_n \in E, T_n \in \mathrm{R}_+, 0 = T_0 \leq T_1 \leq T_2 \cdots$

この確率過程を$(\mathrm{X}, \mathrm{T}) = \{\mathrm{X}_n, T_n; n \in \mathrm{N}_+\}$

確率過程としての交通行動

交通行動が確率過程として記述されるとし、数学的な定式化をし、
頻繁に用いられる確率過程モデルについて述べる

- 再生過程
- マルコフ再生過程
- マルコフ過程
- **マルコフ連鎖**

状態空間 E　非負の整数の集合 N_+

$\mathrm{R}_+ = [0,+\infty]$

n番目の遷移後の状態 $\mathrm{X}_n, n \in \mathrm{N}_+$

n番目の遷移が生じる時刻 T_n

$\mathrm{X}_n \in E, T_n \in \mathrm{R}_+, 0 = T_0 \leq T_1 \leq T_2 \cdots$

この確率過程を$(\mathrm{X}, \mathrm{T}) = \{\mathrm{X}_n, T_n; n \in \mathrm{N}_+\}$

確率過程を(X, T)

離散時点、$S_1, S_2, \ldots, S_K (0 \leq S_1 < S_2 < \ldots < S_K)$で観測されたとし、$S_t$で観測された行動を$Y_t$

$$Y = \{Y_t; t = 1,2, \ldots, K\}$$

ある整数k(>0)について

$$P_r[Y_{n+1} = j | Y_n, Y_{n-1} \ldots, Y_0] = P_r[Y_{n+1} = j | Y_n, Y_{n-1} \ldots, Y_{n-k+1}] \forall \mathrm{j} \in E, n \in \mathrm{N}_+$$

が成立するとき、確率過程Yはk次のマルコフ連鎖と呼ばれる。

通常はk=1で、

$$P_r[Y_{n+1} = j | Y_n = i, Y_{n-1} = i' \ldots, Y_0 = i^0] = P_r[Y_{n+1} = j | Y_n = i] = p_{ij}$$

離散時間パネルデータの解析①

- 誤差要因モデル
- 分布ラグモデル
- ラグ付き内生変数

行動主体をnとし、パネル調査から交通行動の離散時点での観測地が

$$Y = \{Y_{nt}; t = 1,2, \dots, \kappa\}, n = 1, \dots, N$$

と与えられている。

（交通行動は打ち切りなしの実数値）

$$Y_{nt} = \mu + \beta' x_{nt} + \boxed{\varepsilon_{nt}} = \mu + \beta' x_{nt} + \boxed{\alpha_n} + \boxed{\tau_t} + \boxed{u_{nt}} \quad t = 1,2, \dots, \kappa, \quad n = 1, \dots, N$$

x_{nt}:時点S_tでnについて観測された外生変数のベクトル$(\mathrm{K} \times 1)$,

β:$(\mathrm{K} \times 1)$の係数ベクトル　μ:切片

α_nー行動主体間で異なった値、各主体については時点間で変化しない

τ_tー時点間で変化する、各時点において主体間で一定

u_{nt}ー純粋な誤差項。主体間、時点間でi.i.d.

Fixed-effects model(母数モデル) ⟺ α_n τ_t ⟺ Random-effects model(変量モデル)

ダミー変数(定数項)　確率変数

離散時間パネルデータの解析①

- 誤差要因モデル
- 分布ラグモデル
- ラグ付き内生変数

時点S_tでの行動が時点S_{t-R}からS_tに至る間の説明変数の関数

$$Y_{nt} = \boxed{\mu} + \sum_{r=0}^{R} \boxed{\beta'} x_{n,t-r} + \boxed{u_{nt}}$$

$$Y_{nt} = \eta + \beta' x_{nt} + \theta \boxed{Y_{n,t-1}} + w_{nt}$$ ←前観測時点で行動の測定値を説明変数

θ:スカラー　w_{nt}:誤差項　$Y_{n,t-1}$:ラグ付き内生変数

行動過程が開始してR($\in N_+$)時点経過したとき、

$$Y_{nt} = \boxed{\frac{1-\theta^{R+1}}{1-\theta}\eta} + \sum_{r=0}^{R} \boxed{\theta^r \beta'} x_{n,t-r} + \boxed{\sum_{r=0}^{R} \theta^r w_{n,t-r}}$$

離散時間パネルデータの解析②

- Heckmanの動的モデル
- 初期条件
- 状態依存と異質性

$$Y_{nt}^{*} = \beta' x_{nt} + \sum_{l=1}^{t-1} \gamma_l Y_{n,t-l} + \varphi \sum_{s=1}^{t-1} \prod_{l=1}^{s} Y_{n,t-l} + \alpha_n + u_{nt}$$

$$Y_{nt} = \begin{cases} 1, & if\ Y_{nt}^{*} \geq 0 \\ 0, & otherwise \end{cases} \quad n = 1, \dots, N\ t = 1,2, \dots, \kappa$$

状態空間E＝{0,1}

$$Y_{nt}^{*} = \beta' x_{nt} + \alpha_n + u_{nt} \quad Y_{nt} \begin{cases} 1, & if\ Y_{nt}^{*} \geq 0 \\ 0, & otherwise \end{cases} \quad n = 1, \dots, N\ t = 1,2, \dots, \kappa$$ （Y_{nt}^{*}は潜在変数）

誤差要因α_nはx_{nt}から独立で分布G_aを持つ、u_{nt}ー主体間、時点間でi.i.d.累積分布関数F_u

$$Pr[Y_{nt} = 1] = Pr[\ Y_{nt}^{*} \geq 0] = 1 - F_u(-(\beta' x_{nt} + \alpha_n))$$

$$Pr[Y_{nt} = 0] = F_u(-(\beta' x_{nt} + \alpha_n))$$

尤度関数　$\ln L = \sum_{n=1}^{N} \ln \int \prod_{t=1}^{\kappa} \{1 - F_u(-(\beta' x_{nt} + s))\}^{Y_{nt}} \{F_u(-(\beta' x_{nt} + s))\}^{1-Y_{nt}} dG_a(s|\delta)$

を最大化することで、β, δが推定される。

離散時間パネルデータの解析②

- Heckmanの動的モデル
- 初期条件
- 状態依存と異質性

l=1,...,h(≦ κ-1)について$\gamma_l \neq 0$の場合($Y_{n,t-1}$,..., $Y_{n,t-h}$)の一部が$t = 1,2,\ldots$,hについて観測されないことになる。
①初期条件や観測以前の履歴は外生的である。
②行動過程が均衡状態にある。
と仮定する。

$$Y_{nt}^* = \beta' x_{nt} + \gamma_l Y_{n,t-l} + \alpha_n + u_{nt}$$

真の状態依存

ある状態にいたという経験によって知覚、選好、制約条件等の個人の意思決定に影響する要因が変化し、結果として将来の行動が変化

$\gamma_l \neq 0$で$\alpha_n \equiv 0$

虚偽の状態依存

個人がある状態を経験する確率には観測された要因では説明できない差異が個人間であるものの、この確率そのものはその状態を経験するか否かに影響されない

$\gamma_l = 0$で$0 < \sigma_\alpha^2 < \infty$

パネルデータの有効性

~~パネルデータが行動過程を正確に反映したもの~~

<u>離散時点</u>での観測値
からなるパネルデータ

難しい！

<u>連続時間軸上</u>の
行動過程を規定する
諸パラメーターの推定

回顧データを信頼
できる形で収集

パネル調査から
連続データを構築

基礎概念

- 生存時間モデルとは
- 打ち切り
- 生存関数とハザード関数

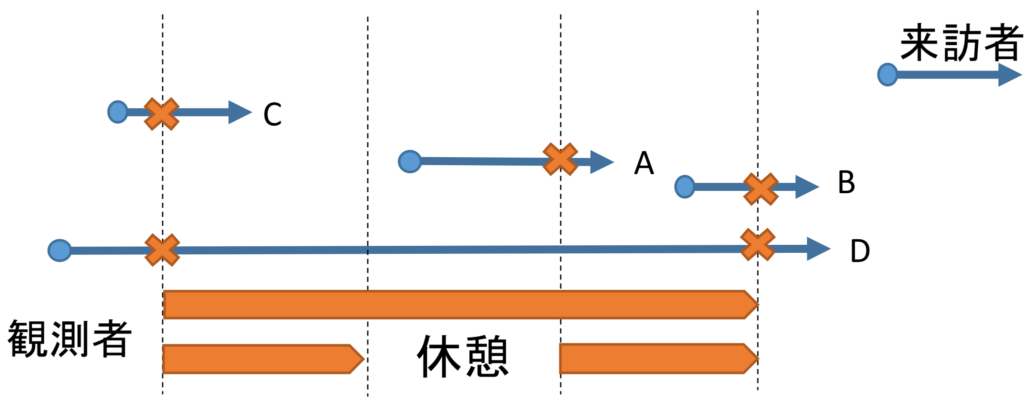
ある基準の時刻からある事象が生起、あるいは終了するまでの時間の解析を対象とするモデル［期間モデル］

特徴

解析対象となる事象が生起するまでの時間は必ず正の値

時間の分布の裾が右に長くなることが多い

分布の正規性を仮定することが適切でない場合が多い



基礎概念

- 生存時間モデルとは
- 打ち切り
- 生存関数とハザード関数

対象とする事象の生起するまでの時間Tの分布を以下のように表すことが多い

生存関数S(t)

ー対象とする事象がある時点tにおいてまだ生起していない確率

ハザード関数h(t)

ー対象とする事象がある時点tにおいてまだ生起していないという条件の下で、次の瞬間に事象が生起するという条件付き確率密度

生存関数S(t)

$$\mathrm{S(t)} = P_r(T \geq t) = 1 - P_r(T \leq t) = 1 - F(\mathrm{t})$$

ハザード関数h(t)

$$h(t) = \lim_{\Delta t \to 0} \frac{P_r(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{S(t) - S(t+\Delta t)}{\Delta t * S(t)} = -\frac{d(\log \mathrm{S(t)})}{dt} = \frac{f(t)}{S(t)}$$

$f(t)$はTの確率密度関数

生存関数S(t)とハザード関数h(t)の関係

$$\mathrm{S}(t) = exp\left(-\int_0^t h(u)du\right)$$

生存時間の解析方法

生存時間のモデル化	共変量がパラメーターとしてモデルに導入されている	生存時間の分布に特定の確率分布が仮定されている	交通行動分析の分野での使用
Non-parametric model	×	×	△
Semi-parametric model	○	×	○
Parametric model	○	○	○

Kaplan-Meier法

d_i＝時点t_iで事象が生起したケース数、事象が生起した時点を$t_1, t_2,\ldots$

n_1＝時点t_iの直前のリスク集合の大きさ(＝その直前に観測されていたケース数)

生存時間がt_{i-1}より長いという条件の下で生存時間がt_i以上である条件付き確率の推定量

$$\frac{\hat{S}(t)}{\hat{S}(t_{i-1})} = \frac{n_i - d_i}{n_i}$$

生存関数S（t）のKaplan-Meier推定量　$\hat{S}(t) = \prod_{t_i<t} \frac{n_i - d_i}{n_i}$

生存時間の解析方法

- 生存時間のモデル化
- Non-parametric model
- Semi-parametric model
- Parametric model

	共変量がパラメーターとしてモデルに導入されている	生存時間の分布に特定の確率分布が仮定されている	交通行動分析の分野での使用
Non-parametric model	×	×	△
Semi-parametric model	○	×	○
Parametric model	○	○	○

比例ハザードモデルでのハザード関数

$$h(t|x_i) = h_0(t)\exp(-\beta x_i)$$

$h(t|x_i)$:共変量ベクトルx_iを持つケースiのハザード関数、$h_0(t)$:基準ハザード、

β:未知パラメーターベクトル、x_i:ケースiの共変量ベクトル

(ケース間でのハザード関数の比は時点によらず一定と仮定)

パラメーターの推定:部分尤度法、Han and Hausman法

生存時間の解析方法

• 生存時間のモデル化	共変量がパラメーターとしてモデルに導入されている	生存時間の分布に特定の確率分布が仮定されている	交通行動分析の分野での使用
• Non-parametric model	×	×	△
• Semi-parametric model	○	×	○
• Parametric model	○	○	○

①比例ハザードモデル

②加速故障モデル

$$h(t|x_i) = h_0(t\exp(-\beta x_i))\exp(-\beta x_i)$$

生存時間への共変量の

影響の仕方が異なる

→2つのモデルのうち対象とする

問題に適したモデルを選択する

確率分布	ハザード関数$h_0(t)$	生存関数$S_0(t)$
指数分布	λ	exp(−λt)
ワイブル分布	$\gamma\lambda_t^{\gamma-1}$	exp(−λ_t^{γ})
対数ロジスティック分布	$\frac{\gamma\lambda_t^{\gamma-1}}{1+\lambda_t^{\gamma}}$	$\frac{1}{1+\lambda_t^{\gamma}}$
対数正規分布	$\frac{1}{t}\varphi\left(\frac{\log(t)-\mu}{\sigma}\right)\Big/ 1-\varphi\left(\frac{\log(t)-\mu}{\sigma}\right)$	$1-\varphi\left(\frac{\log(t)-\mu}{\sigma}\right)$
ゴンペルツ分布	γexp(−λt)	exp((γ/λ)(exp(−λt)−1))

基本モデルの拡張

- 非観測異質性
- 時間依存性共変量
- 競合危険

Parametric model：
モデルの共変量以外の異質性は存在しない

確率分布の未知パラメーターの推定値、
共変量の係数ベクトルの推定値
がバイアスを受ける

非観測異質性を考慮したモデルの推定法

例：ワイブル分布

$$S(t) = \left\{1 + \theta\lambda t^{\gamma}\right\}^{-1/\theta}$$

通常：$S(t) = \exp(-\lambda t^{\gamma})$

基本モデルの拡張

- 非観測異質性
- 時間依存性共変量
- 競合危険

比例ハザードモデル：
時間が経過しても共変量が
変化しない

対象とする事象が生起する原因が複数考えられる場合、
観測によっていずれの原因によって個々の事象が生起したかを
特定することが可能であれば、それらを区別して分析を行うことが可能

各原因が独立

$$h(t)=\sum_{m=1}^{M}h_m(t)\quad S(t)=\prod_{m=1}^{M}S_m(t)$$

各原因が非独立

誤差項の相関をモデルに
導入した上で全ての原因の
パラメーターを同時に推定

適用事例

- 活動時間、滞在時間の分析
- 自動車保有行動の分析

交通需要の時間的な変動←重要

個々人の日常的な自由活動の活動時間

観光地における滞在時間

例）藤井ら、1997

個人の1日の活動を再現するためのシミュレーターの1要素として
活動時間を再現する生存時間モデルが適用

活動開始時間を基準として活動終了時刻までの時間(活動時間)を、
ワイブル分布を仮定したParametric modelが適用

適用事例

- 活動時間、滞在時間の分析
- 自動車保有行動の分析

時間軸上における世帯の自動車取り替え更新行動の分析

←世帯の自動車保有率や自動車保有台数に関する研究

(モビリティを説明する要因)

自動車保有期間・自動車取り替え行動の分析

例）山本俊行、木村誠司、北村隆一：取替更新行動間の相互影響を考慮した世帯の自動車取替更新行動モデルの構築、土木計画学研究・論文集、No15、pp593-599、1997

取替更新行動に、買い替え・追加購入・破棄の3種類が存在するとし、競合危険モデルを構築し、n台保有世帯の場合2n+1の原因によって更新行動が生起するとし、モデル化している。

Parametricな比例ハザードモデルにワイブル分布を適用した定式化を行っている。

まとめ

- 動的モデル

  交通行動の動的特性

  確率過程としての交通行動

  離散時間パネルデータの解析①②

  パネルデータの有効性

- 生存時間モデル

  基礎概念

  生存時間の解析方法

  基本モデルの拡張

  適用事例