

Statical models of discrete choice processes

Rust, John(1988)

Transportation Research Part B: Methodological 22.2:125-158

交通研究室 B4

倉澤龍平

Discrete choice process

離散選択過程（Discrete choice process）とは

各時点で有限の選択肢の中から選ぶような行動を記述する確率過程

本論文の目的

離散選択過程の定式化/推定に関する研究成果を紹介し、交通分野にどのように適用できるのかを示す

※論文発表は1988年。それまでの結果をまとめたものとなる。

「今までの研究成果をまとめた」という体の論文なので、普通の論文とはちょっと勝手が違うかもしれません。

目次

1. Introduction

- 構造モデルと誘導モデルの違い

2. Reduced-form Models

- 誘導モデルに関するトピックについて説明

3. Structural Models

- 構造モデルに関するトピックについて， Rustの研究成果を中心に説明

論文全体として量が多いので，内容を絞って

- 構造モデルと誘導モデルの違いを理解する
- 構造推定に使われる計算手法をつかむ

ことを目標として，発表を行いたいと思います。

1. Introduction

構造モデルと誘導モデル

構造モデル (Structural model) とは

- 確率過程のパラメータが、動学最適化の解として求まるもの
- 「確率過程のパラメータ」と「エージェントの目的関数のパラメータ」を別々に推定する

構造モデルと誘導モデル

誘導モデル (Reduced-form model) とは

- 確率過程のパラメータを、背景にある最適化問題から明示的に導出しない方法
- 「確率過程のパラメータ」と「エージェントの目的関数のパラメータ」を分離せず、結合して扱う。

誘導モデルのメリット・デメリット

メリット

- 複雑な情報を少数のパラメータに集約して扱うことができる。
- 推定に比較的便利で、データの相互関係を正確に反映できる関数形を指定できる。

デメリット

- 政策による影響を正しく見積もれないことがある。
 - 政策が変わることで「確率過程のパラメータ」が変更される。
それにより、誘導モデルのパラメータも変更されるから。

構造モデルのメリット・デメリット

メリット

- 構造を推定することで、政策の変化に対応したパラメータを得ることができる

デメリット

- 閉形式の解を得るのが難しい
 - 閉形式の解が得られるのは極めて厳しい仮定をしたときのみで、それなら誘導モデルを使ったほうがいい
- データと矛盾するような関係式を事前に構築してしまうと、データ同士の関係がわからなくなり、誤った推定をしてしまう。

2. Reduced-Form models

一般論における文字設定

M エージェントの数

T_m : エージェント m のデータ収集期間

x_t^m : 時刻 t に観測したエージェント m の説明変数 (K 次元ベクトル)

i_t^m : 時刻 t に観測したエージェント m の行動選択

$C(x_t)$: 説明変数 x_t が与えられたときの選択肢集合

最尤法による推定

$l(i_1, \dots, i_T, x_1, \dots, x_T; \theta) : i, x$ の同時分布

最尤推定量 $\hat{\theta}$ は以下の式により定義される.

$$\hat{\theta} = \operatorname{argmax} L(\theta) \equiv \sum_{m=1}^M \log l(i_1, \dots, i_T, x_1, \dots, x_T; \theta)$$

また, 条件付き分布 $l(i_1, \dots, i_T, |x_1, \dots, x_T; \theta)$ すなわち,

$$l(i_1, \dots, i_T, |x_1, \dots, x_T; \theta) = \frac{l(i_1, \dots, i_T, x_1, \dots, x_T; \theta)}{\sum_{i_1} \cdots \sum_{i_T} l(i_1, \dots, i_T, x_1, \dots, x_T; \theta)}$$

の形を事前に指定した上で, 条件付き尤度

$$L^c(\theta) \equiv \sum_{m=1}^M \log l(i_1, \dots, i_T, |x_1, \dots, x_T; \theta)$$

の最大化を目指すこともある.

具体例1：ベルヌーイ過程

説明変数がなく、2つの選択肢から選択する（つまり、 $C(x_t) = \{0, 1\}$ ）ような状況を考える。パラメータを $\theta = \Pr(i_t = 1)$ とする。このとき、

$$l(i_1, \dots, i_T; \theta) = \prod_{t=1}^T \theta^{i_t} (1 - \theta)^{1-i_t}$$

このときのパラメータ θ は選択確率を表しているが、この選択の背景にあるメカニズムは全く無視されてしまっている。

→状況が変わってしまうと、パラメータ θ の値は変わってしまうので、政策の変更に
対応できない

具体例1：ベルヌーイ過程

説明変数がなく，2つの選択肢から選択する（つまり， $C(x_t) = \{0, 1\}$ ）ような状況を考える．パラメータを $\theta = \Pr(i_t = 1)$ とする．このとき，

$$l(i_1, \dots, i_T; \theta) = \prod_{t=1}^T \theta^{i_t} (1 - \theta)^{1-i_t}$$

ここから最尤推定を行うと，

$$\hat{\theta} = \sum_{t=1}^T \frac{i_t}{T}$$

複数人の場合でも，全員が同じパラメータで行動すると考えれば，

$$\hat{\theta} = \frac{\sum_{m=1}^M \sum_{t=1}^{T_m} i_t^m}{\sum_{m=1}^M T_m}$$

具体例1：ベルヌーイ過程

「複数人の場合でも、全員が同じパラメータで行動すると考えれば」 ← そんなわけない

⇒ 異質性の考慮のため、選択確率 p がある **確率分布 $h(p; \theta)$ に従う確率変数** と考える

このときの尤度関数は、

$$L(\theta) = \sum_{m=1}^M \log \left\{ \int_0^1 \prod_{t=1}^{T_m} p^{i_t^m} (1-p)^{(1-i_t^m)} h(p, \theta) dp \right\}$$

p は確率的に分布するので、期待値を出すことで尤度関数としている。

例えば、分布としてベータ分布

$$h(p, \theta) = \frac{1}{B(\theta_1, \theta_2)} p^{\theta_1-1} (1-p)^{\theta_2-1}$$

を仮定すると、尤度関数は、

$$L(\theta) = \sum_{m=1}^M \log \left\{ \frac{B(\theta_1 + S_m, \theta_2 + T_m - S_m)}{B(\theta_1, \theta_2)} \right\} \quad \text{where } S_m = \sum_{t=1}^{T_m} i_t^m$$

具体例2：多項過程

- ベルヌーイ過程の選択肢をN個にしたver.
- 選択肢は $C = \{1, \dots, N\}$ でパラメータは $\theta = \{\theta_1, \dots, \theta_{N-1}\}$
- そして, $i_t(k) = 1$ if $i_t = k$ とする.
- 先程と同じように, それぞれの確率が分布 $\mathbf{h}(\mathbf{p}; \boldsymbol{\theta})$ に従うとする. この分布として, ベータ分布を拡張した, Dirichlet分布を用いる.

$$h(\mathbf{p}; \boldsymbol{\theta}) = \frac{\Gamma(\theta_1 + \dots + \theta_N)}{\Gamma(\theta_1) \dots \Gamma(\theta_N)} \prod_{i=1}^N p_i^{\theta_i - 1}, \quad \text{where } 0 < p_i, \sum_{i=1}^N p_i = 1$$

先程と同じような計算をすることで, 尤度関数の式が得られる.

$$L(\boldsymbol{\theta}) = \sum_{n=1}^M \log \left\{ \frac{\Gamma\left(\sum_{i=1}^N \theta_i\right)}{\Gamma\left(\sum_{i=1}^N (\theta_i + S_i^m)\right)} \frac{\prod_{i=1}^N \Gamma(\theta_i + S_i^m)}{\prod_{i=1}^N \Gamma(\theta_i)} \right\} \quad \text{where } S_k^m = \sum_{t=1}^{T_m} i_t^m(k)$$

単純な二項過程・多項過程の失敗

- 二項過程・多項過程を用いた推定はいずれもあまり良い結果が出ていない。
(Frank(1962), Bass and Wright(1980)など)
- これは
 1. 選択行動はもっと複雑な枠組みが必要であること
 2. 選択行動には時間的な関連があること（繰り返しの購買など）を示唆している。
- 行動選択に時間的な関連を入れ込む方法を、ここから2つ紹介する。

対策①：Conditional multinomial logit model

- 先程の多項過程を，時間に関する説明変数を入れ込めるように拡張し，尤度関数を以下のように定めたもの.

$$l(i_1, \dots, i_T, |x_1, \dots, x_T; \theta) = \prod_{t=1}^T \frac{\exp\{u(x_t, i_t, \theta)\}}{\sum_{j \in C(x_t)} \exp\{u(x_t, j, \theta)\}}$$

- u の式は説明変数とパラメータの線形和とする事が多い

$$u(x_t, j, \theta) = u(x_t, j)\theta$$

- 時間ごとに変わる変数を用いて作ったモデルであり，選択肢同士の直接的な関連を表現するものではない.

対策②：Markov chains

- 未来の挙動が現在の値だけで決定される確率過程
- 最もシンプルなのは，以下のような形（時刻 t と $t+1$ の行動の間に関連がある）

$$l(i_1, \dots, i_T; \theta) = \prod_{t=1}^T \theta_{i_t, i_{t+1}}$$

- マルコフ連鎖の場合，異質性を考慮するのはより大変になる
 - 推定すべき遷移確率が行列の形になるため，個人ごとの遷移行列の推定，もしくは遷移行列の分布の推定が必要になる

マルコフ連鎖における異質性の考慮

- 遷移確率 p_{ij} を説明変数 x を用いた関数として表す.

$$l(i_1, \dots, i_T | i_0, x_1, \dots, x_T; \theta) = \prod_{t=1}^T p_{i_{t-1}, i_t}(x_t, \theta)$$

- 例えば, Manski and Sherman(1981) やBerkovec and Rust(1985) では, 自動車保有モデルにおいて, 以下のような説明変数を加えたところ, 係数が大きな正の数となり, 尤度が上昇したことを確認した.

i_t : t 期に自動車を保有していたか否かを表す2値データ

$$d(i_t, i_{t-1}) = \begin{cases} 1 & i_t = i_{t-1} \text{ のとき} \\ 0 & \text{otherwise} \end{cases}$$

- 車を保有している人は保有したままで, 保有していない人は保有していないままでいる傾向があると考えられるともっともらしい.

3. Structural models

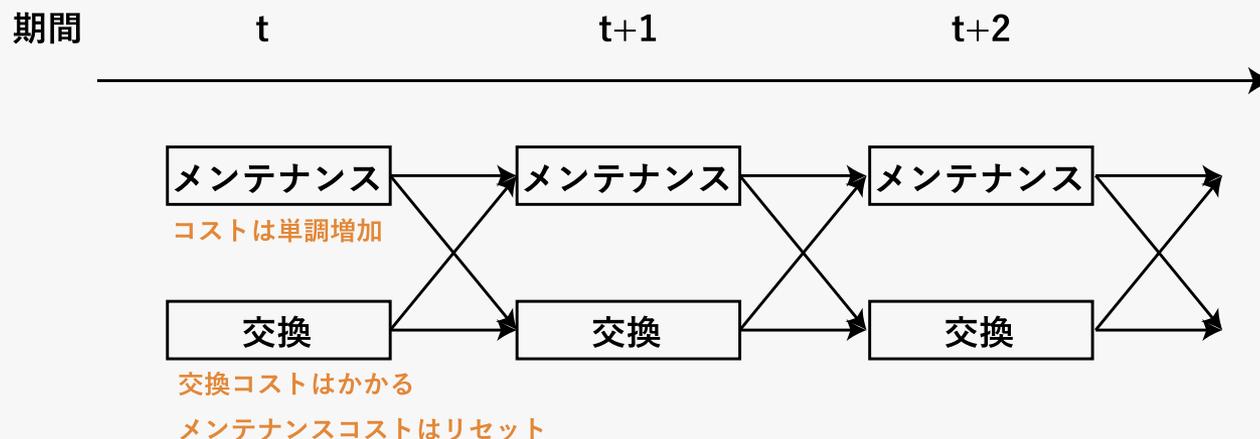
構造モデルについて

構造モデル (Structural model) とは (再掲)

- 確率過程のパラメータが、動学最適化の解として求まるもの
- 「確率過程のパラメータ」と「エージェントの目的関数のパラメータ」を別々に推定する

まずは具体例

- 「バスのエンジンを交換するのに適切な時期はいつか？」という問題を動的計画法の枠組みとして解くことを考える。
- ざっくりいうと，
 - 每期，バスのエンジンを通常メンテナンスで済ますか，交換するかを選択する
 - バスエンジンのメンテナンスコストは，長く使えば使うほど増えていくはずである。
 - しかしバスのエンジンを買い換えるのにも費用がかかる→このような状況の中で，いつエンジンを買い換えるのがよいか？



文字設定と仮定

文字設定

表 1 バスエンジン買い替え問題の文字設定

文字	意味
x_t	期間 t におけるバスエンジンの累積の走行距離 (エンジンが買い換えられるとリセット)
$c(x_t, \theta_1)$	期間 t におけるバスのオペレーションコスト (メンテナンスコスト, バスが故障した場合の乗客へ与える被害など)
P	古いエンジンを処理するのにかかる費用
\bar{P}	新しいバスのエンジンを導入するのにかかる費用
i_t	期間 t に下した判断 $i_t = 0$ ならエンジンを使い続ける, $i_t = 1$ なら買い換える
β	時間割引率

仮定

- $c(x_t, \theta_1)$ は単調増加で, 微分可能な関数
- t 期の走行距離 (x_t のことではない) はパラメータ θ_2 の指数分布に従う.

価値関数の定式化

期間 t に得られる効用を u_t とすると,

$$u(x_t, i_t, \theta_1) = \begin{cases} \text{メンテナンスコスト} \\ -c(x_t, \theta_1) & i_t = 0 \text{ のとき} \\ -(\bar{P} - P + c(0, \theta_1)) & i_t = 1 \text{ のとき} \\ \text{交換コスト} \end{cases}$$

この θ_1 が「目的関数のパラメータ」

であり, 価値関数は,

$$V_\theta(x_t) = \sup_{\Pi} E \left\{ \sum_{j=t}^{\infty} \beta^{j-t} u(x_j, f_j, \theta_1) \mid x_t \right\}$$

将来まで考えたときの効用の和

表1 バスエンジン買い替え問題の文字設定

ただし, f_j は i_t の値を決める方策を表す関数で,

$$i_t = f_t(x_t, i_{t-1}, x_{t-1}, i_{t-2}, x_{t-2}, \dots)$$

となる. Π は方策集合で,

$$\Pi = \{f_t, f_{t+1}, \dots\}$$

文字	意味
x_t	期間 t におけるバスエンジンの累積の走行距離 (エンジンが買い換えられるとリセット)
$c(x_t, \theta_1)$	期間 t におけるバスのオペレーションコスト (メンテナンスコスト, バスが故障した場合の乗客へ与える被害など)
P	古いエンジンを処理するのにかかる費用
\bar{P}	新しいバスのエンジンを導入するのにかかる費用
i_t	期間 t に下した判断 $i_t = 0$ ならエンジンを使い続ける, $i_t = 1$ なら買い換える
β	時間割引率

走行距離の分布

「 t 期の走行距離(x_t ではない!)はパラメータ θ_2 の指数分布に従う」という仮定から、走行距離の確率密度関数は以下ようになる。

$$p(x_{t+1}|x_t, i_t, \theta_2) = \begin{cases} \theta_2 \exp\{-\theta_2(x_{t+1} - x_t)\} & \text{メンテナンスで済ませた} \\ \theta_2 \exp\{-\theta_2(x_{t+1})\} & \text{交換を行った} \\ 0 & \text{otherwise} \end{cases}$$

(Note: The conditions for the cases are $i_t = 0$ and $x_{t+1} \geq x_t$ for the first case, and $i_t = 1$ and $x_{t+1} \geq 0$ for the second case.)

この θ_2 が「確率過程のパラメータ」

Bellman's equationの利用

Bellman方程式より，以下の式が成り立つ

$$V_{\theta}(x_t) = \max_{i_t} [\underbrace{u(x_t, i_t, \theta_1)}_{\text{今期の効用}} + \beta \underbrace{EV_{\theta}(x_t, i_t)}_{\text{次期の価値関数の期待値}}]$$

ただし，

$$EV_{\theta}(x_t, i_t) = \int_{-\infty}^{\infty} \underbrace{V_{\theta}(y)}_{\text{次期の価値関数}} \underbrace{p(y|x_t, i_t, \theta_2)}_{\text{次期状態の確率分布}} dy$$

以上のことを用いると，最適方策がわかる（詳細はRust(1985a)を参照）

$$i_t = f(x_t, \theta) = \begin{cases} 1 & x_t > \gamma(\theta_1, \theta_2) \text{ のとき} \\ 0 & x_t \leq \gamma(\theta_1, \theta_2) \text{ のとき} \end{cases}$$

ただし， γ は以下の式の解である．

$$(\bar{P} - \underline{P})(1 - \beta) = \int_0^{\gamma(\theta_1, \theta_2)} [1 - \beta \exp(-\theta_2(1 - \beta)y) \frac{\partial c(y, \theta_1)}{\partial y}] dy$$

パラメータの決定手法

- 観測できるデータは M 台のバスの「 t 期に走った距離」「 t 期にエンジンを入れ替えたか否か」である。つまり、 $(i_1^m, \dots, i_{T_m}^m, x_1^m, \dots, x_{T_m}^m)$
- さらに、 $\bar{P} - \underline{P}$ と β に関しては観測可能とする。これで未知パラメータは (θ_1, θ_2) のみのなった。
- 最尤法によりパラメータを求める。そのため、まず1台のバスに注目して、 $l(i_1, \dots, i_T, x_1, \dots, x_T; \theta)$ を求める。
- j 番目のエンジンの寿命を T_j とおく。このとき、各値は以下の表のように関連している。

t	1	...	T_1	$T_1 + 1$...	$T_1 + T_2$...	T
i_t	0	...	<u>1</u>	0	...	1	...	
x_t	x_1	...	交換 over γ		...	over γ	...	

パラメータの決定手法

- このとき、 $T_j - 1$ はパラメータ $\theta_2\gamma$ のポアソン過程となる（詳細はRust(1985b)) .

$$\Pr(\bar{T}_i = t) = \frac{[\theta_2\gamma]^{t-1} \exp\{-\theta_2\gamma\}}{(t-1)!}$$

エンジンの寿命が t となる確率

- これを用いて、 $l(i_1, \dots, i_T, x_1, \dots, x_T; \theta)$ を求める.

まず、1個目のエンジンを交換するまでの分布、すなわち

$$l(i_1, \dots, i_{T_1}, x_1, \dots, x_{T_1}; \theta)$$

を知りたい。さらにこれを、

$$l(i_1, \dots, i_{T_1}, x_1, \dots, x_{T_1}; \theta) = l(x_1, \dots, x_{T_1}, |i_1, \dots, i_{T_1}, \theta)l(i_1, \dots, i_{T_1}; \theta)$$

と分割して、それぞれ求める。

パラメータの決定手法

- $l(i_1, \dots, i_{T_1}; \theta)$ は, T_1 でエンジンの交換を行ったことを表しているので,

$$l(i_1, \dots, i_{T_1}; \theta) = \Pr(\bar{T}_i = T_1) = \frac{[\theta_2 \gamma]^{T_1-1} \exp\{-\theta_2 \gamma\}}{(T_1 - 1)!}$$

0,0,...,0,1となっている

- また, 走行距離は指数分布に従うことから,

$$\begin{aligned} l(x_1, \dots, x_{T_1} | i_1, \dots, i_{T_1}; \theta) &= \prod_{t=1}^{T_1} \theta_2 \exp\{-\theta_2(x_t - x_{t-1})\} \\ &= \theta_2^{T_1} \exp\{-\theta_2 x_{T_1}\} \end{aligned}$$

必要なのは総走行距離だけ!

- 以上より, 1個目のエンジンを交換するまでの尤度は,

$$l(i_1, \dots, i_{T_1}, x_1, \dots, x_{T_1}; \theta) = \theta_2^{T_1} \exp\{-\theta_2 x_{T_1}\} \frac{[\theta_2 \gamma]^{T_1-1} \exp\{-\theta_2 \gamma\}}{(T_1 - 1)!}$$

パラメータの決定手法

- よって、期間 T まで考えると、同時分布は、

$$\begin{aligned}
 & l(i_1, \dots, i_{T_1}, x_1, \dots, x_{T_1}; \theta) \\
 &= \prod_{j=1}^N \theta_2^{T_j} \exp\{-\theta_2 x_{T_j}\} \frac{[\theta_2 \gamma]^{T_j-1} \exp\{-\theta_2 \gamma\}}{(T_j - 1)!} \quad \begin{array}{l} \text{if } i_T = 1 \\ \text{最後に交換をした} \end{array} \\
 & \prod_{j=1}^{N-1} \theta_2^{T_j} \exp\{-\theta_2 x_{T_j}\} \frac{[\theta_2 \gamma]^{T_j-1} \exp\{-\theta_2 \gamma\}}{(T_j - 1)!} \\
 & \times \theta_2^{T_N} \exp\{-\theta_2 x_{T_N}\} \left\{ \frac{1 - \sum_{t=1}^{T_N} \frac{[\theta_2 \gamma]^{t-1} \exp\{-\theta_2 \gamma\}}{(t-1)!}}{\theta_2 \gamma} \right\} \quad \begin{array}{l} \text{if } i_T = 0 \\ \text{最後に交換をしなかった} \end{array} \\
 & \qquad \qquad \qquad \text{1} \sim T_N \text{の間で交換をしなかった}
 \end{aligned}$$

- これは1台のバスのみのデータなので、 M 台のバスに関してこれをかけてあげること
とで、尤度関数を得る。

パラメータの決定手法

$$\begin{aligned}
 & l(i_1, \dots, i_{T_1}, x_1, \dots, x_{T_1}; \theta) \\
 &= \prod_{j=1}^N \theta_2^{T_j} \exp\{-\theta_2 x_{T_j}\} \frac{[\theta_2 \gamma]^{T_j-1} \exp\{-\theta_2 \gamma\}}{(T_j - 1)!} && \text{if } i_T = 1 \\
 & \prod_{j=1}^{N-1} \theta_2^{T_j} \exp\{-\theta_2 x_{T_j}\} \frac{[\theta_2 \gamma]^{T_j-1} \exp\{-\theta_2 \gamma\}}{(T_j - 1)!} \\
 & \times \theta_2^{T_N} \exp\{-\theta_2 x_{T_N}\} \left\{ 1 - \sum_{t=1}^{T_N} \frac{[\theta_2 \gamma]^{t-1} \exp\{-\theta_2 \gamma\}}{(t-1)!} \right\} && \text{if } i_T = 0
 \end{aligned}$$

- この式は (θ_1, θ_2) の関数だが、 θ_1 は γ という形で間接的にしか登場していない。
 → まず γ と θ_2 を最尤推定し、そこから θ_1 を

$$(\bar{P} - \underline{P})(1 - \beta) = \int_0^{\gamma(\theta_1, \theta_2)} [1 - \beta \exp(-\theta_2(1 - \beta)y) \frac{\partial c(y, \theta_1)}{\partial y}]$$

を用いて求める。例えば、 $c(y, \theta_1) = \theta_1 y$ とおくと、

$$\hat{\theta}_1 = \frac{(\bar{P} - \underline{P})(1 - \beta)}{\int_0^{\gamma} [1 - \beta \exp\{-\hat{\theta}_2(1 - \beta)y\}] dy}$$

構造モデルの確認

構造モデルは、誘導モデルと違い、

- 効用関数、遷移確率といったプリミティブな部分のパラメータを推定している。
- 尤度関数が $\{i_t, x_t\}$ の確率密度から得られる

といった特徴があることが確認できた。

閉形式解の限界

閉形式解（Closed-form solution）とは

解が初等関数の計算だけで求まるもの。

今回の例では閉形式の解を求めることができた。しかし、実際に閉形式の解が得られることは極めて少ない。

また、閉形式の解が得られた場合というのは、往々にして

- 閉形式の解を得るための**恣意的な数学的仮定が置かれる**
- 閉形式の解を得るために問題を**過度に単純化している**

といったことが行われているため、モデルのちょっとした変更に対応できなくなってしまう。

閉形式解の限界

今回の例でも

- 閉形式の解を得るための恣意的な数学的仮定が置かれる
- 閉形式の解を得るために問題を過度に単純化している

といったことが行われている。

例えば,

- 「 t 期の走行距離(x_t のことではない)はパラメータ θ_2 の指数分布に従う。」という仮定から、**走行距離の平均と標準偏差が等しくなる**ことが求められるが、実際にはそのようなことは起こっていない。(恣意的な数学的仮定)
- 交換するかの判断基準が x_t という1つの指標でしか評価されていない。実際には他にも様々な変数が有り、それに基づいて総合的に判断をしているはず。(過度な単純化)

誤差項の導入

非観測変数の存在を考慮するために、誤差項を導入する。

$$i_t = f(x_t, \theta) + \epsilon_t$$

このようにすることで、エージェントが最適解から離れた行動をすることを許容できる。

最尤推定アルゴリズム

Rust(1987a)は

- 解が閉形式とならない
- 誤差項を含む

場合の最尤推定アルゴリズムを開発した。 それについて解説する。

問題の定式化

文字設定

文字	意味
$C(x_t)$	制御変数 i_t の有限な選択肢集合
$\epsilon_t = \{\epsilon_t(i) i \in C(x_t)\}$	$C(x_t)$ 次元の誤差項 (非観測変数) ベクトル
$x_t = \{x_t(1), \dots, x_t(K)\}$	K 次元の観測された状態変数のベクトル
$u(x_t, i, \theta_1) + \epsilon_t(i)$	選択肢 i を選んだときの瞬間的な効用
$p(x_{t+1}, \epsilon_{t+1} x_t, \epsilon_t, i_t, \theta_2, \theta_3)$	選択肢 i_t を選んだときの推移確率分布
$\theta = (\beta, \theta_1, \theta_2, \theta_3)$	$1 + K_1 + K_2 + K_3$ 次元の, 推定するパラメータ
$f_t(x_t, \epsilon_t, \theta)$	i_t の値を決める方策
$\Pi = \{f_t, f_{t+1}, f_{t+2}, \dots\}$	方策の集合

問題の定式化

定式化

価値関数は、先ほどと同じようにして、以下のようになる。

$$V_{\theta}(x_t, \epsilon_t) = \sup_{\Pi} E \left\{ \sum_{j=t}^{\infty} \beta^{(j-t)} [u(x_j, f_j, \theta_1) + \epsilon_j(f_j)] \mid x_t, \epsilon_t, \theta_2, \theta_3 \right\}$$

Bellman方程式より、 V_{θ} は以下の式の解となる。

$$V_{\theta}(x_t, \epsilon_t) = \max_{i \in C(x_t)} [u(x_t, i, \theta_1) + \epsilon_t(i) + \beta EV_{\theta}(x_t, \epsilon_t, i)]$$

誤差項の混入がさっきとの違い

ただし、

$$EV_{\theta}(x_t, \epsilon_t, i) = \int_y \int_{\eta} V_{\theta}(y, \eta) p(y, \eta \mid x_t, \epsilon_t, i, \theta_2, \theta_3) dy d\eta$$

よって、方策については以下の式が成り立つ。

$$i_t = f(x_t, \epsilon_t, \theta) = \operatorname{argmax}_{i \in C(x_t)} [u(x_t, i, \theta_1) + \epsilon_t(i) + \beta EV_{\theta}(x_t, \epsilon_t, i)]$$

条件付き独立性の仮定 (CI)

問題を簡単にするために、以下のような仮定を導入する。

$$p(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t, i, \theta_2, \theta_3) = q(\epsilon_{t+1} | x_{t+1}, \theta_2) p(x_{t+1} | x_t, i, \theta_3)$$

仮定の解釈

1. x_{t+1} は ϵ_{t+1} の十分統計量である。
(= ϵ_t と ϵ_{t+1} の関係性は、 x_{t+1} と ϵ_{t+1} の関係性を考えることで一緒に伝わる)
2. x_{t+1} は x_t には依存しているが、 ϵ_t には依存していない

関係性のイメージ図

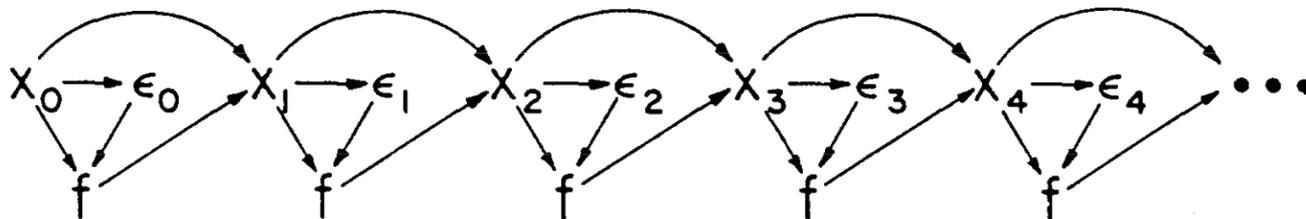


Fig. 1. Pattern of dependence in controlled Markov Process implied by conditional independence assumption (CI).

条件付き独立性の仮定

$$p(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t, i, \theta_2, \theta_3) = q(\epsilon_{t+1} | x_{t+1}, \theta_2) p(x_{t+1} | x_t, i, \theta_3)$$

条件付き独立性の仮定をおいたことにより，以下の定理が成り立つ（詳細はRust(1987a)）

定理1

条件付き独立の仮定が成り立つとする．このとき，

$$G([u(x, \theta_1) + \beta EV_\theta(x)] | x, \theta_2) = \int_\epsilon \max_{j \in C(x)} [u(x, j, \theta_1) + \beta EV_\theta(x, j)] q(\epsilon | x, \theta_2) d\epsilon$$

と G を定めると，各行動の選択確率は，

$$P(i | x, \theta) = G_i([u(x, \theta_1) + \beta EV_\theta(y)] | y, \theta_2)$$

ただし， G_i は G の $u(x, i, \theta_1)$ に関する偏微分を表す．

そしてこのとき， EV_θ は以下の式を満たす．

$$EV_\theta(x, i) = \int_y G([u(y, \theta_1) + \beta EV_\theta(y)] | y, \theta_2) p(y | x, i, \theta_3) dy$$

定理1の解釈

静的な問題と同じやり方で動的な問題が解けるようになるのがポイント。

例えば q の分布として、多変量ガンベル分布を仮定する。

$$q(\epsilon|x, \theta_2) = \prod_{j \in C} \exp(-\epsilon(j) + \theta_2) \exp(-\exp(-\epsilon(j) + \theta_2))$$

このとき、 G は

$$G([u(x, \theta_1) + \beta EV_\theta(x)]|x, \theta_2) = \ln \left\{ \sum_{j \in C(x)} \exp[u(x, j, \theta_1) + \beta EV_\theta(x, j)] \right\}$$

となり、選択確率は、

$$P(i|x, \theta) = \frac{\exp\{u(x, i, \theta_1) + \beta EV_\theta(x, i)\}}{\sum_{j \in C(x)} \exp\{u(x, j, \theta_1) + \beta EV_\theta(x, j)\}}$$

MNLに将来効用を足した形に！

このとき、 EV_θ は以下の式の解となる。

$$EV_\theta(x, i) = \int_y \ln \left\{ \sum_{j \in C(y)} \exp\{u(x, j, \theta_1) + \beta EV_\theta(x, j)\} \right\} p(y|x, i, \theta_3) dy$$

→ EV_θ をもとめることができれば選択確率が求まる！！

尤度関数を求める

選択確率は求められたので，ここから最尤法を用いてパラメータを推定したい。

定理2

条件付き独立の仮定が成り立つとする。このとき，尤度関数は，

$$L^f(x_1, \dots, x_T, i_1, \dots, i_T | x_0, i_0, \theta) = \prod_{t=1}^T P(i_t | x_t, \theta) p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

で表される

これまでのまとめ

条件付き独立性の仮定

$$p(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t, i, \theta_2, \theta_3) = q(\epsilon_{t+1} | x_{t+1}, \theta_2) p(x_{t+1} | x_t, i, \theta_3)$$

を仮定することで、パラメータを以下のようにして推定することができる。

1. 以下の式を解いて EV_θ を求める

$$EV_\theta(x, i) = \int_y \ln \left\{ \sum_{j \in C(y)} \exp\{u(x, j, \theta_1) + \beta EV_\theta(x, j)\} \right\} p(y|x, i, \theta_3) dy$$

2. 選択確率を、以下の式から求める。

$$P(i|x, \theta) = \frac{\exp\{u(x, i, \theta_1) + \beta EV_\theta(x, i)\}}{\sum_{j \in C(x)} \exp\{u(x, j, \theta_1) + \beta EV_\theta(x, j)\}}$$

3. 尤度関数を以下の式で求めて、最尤法を実行

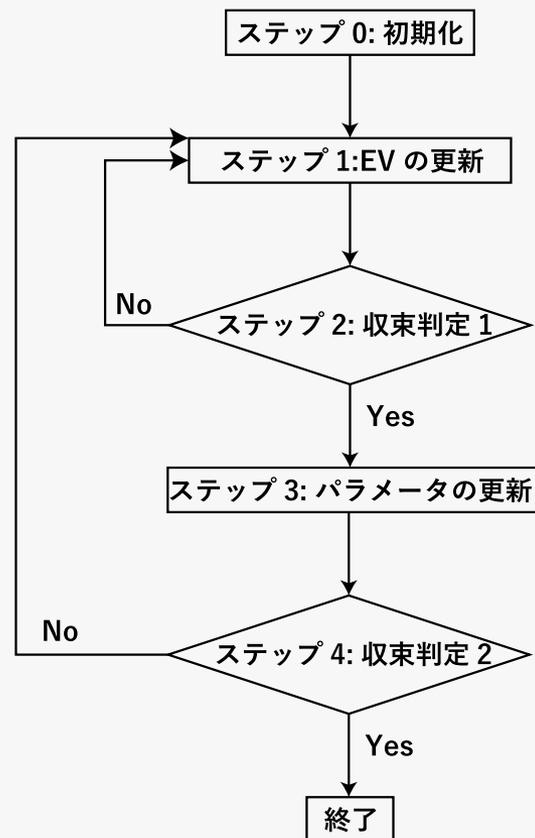
$$L^f(x_1, \dots, x_T, i_1, \dots, i_T | x_0, i_0, \theta) = \prod_{t=1}^T P(i_t | x_t, \theta) p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

NFXPアルゴリズム

この計算を行うためにRust(1987d)が開発したアルゴリズムを**NFXPアルゴリズム** (nested fixed point estimation algorithm)という。

このアルゴリズムは、ざっくりと

1. 不動点問題を解くステップ
 2. 最尤推定をして、パラメータを求めるステップ
- に分けられる。

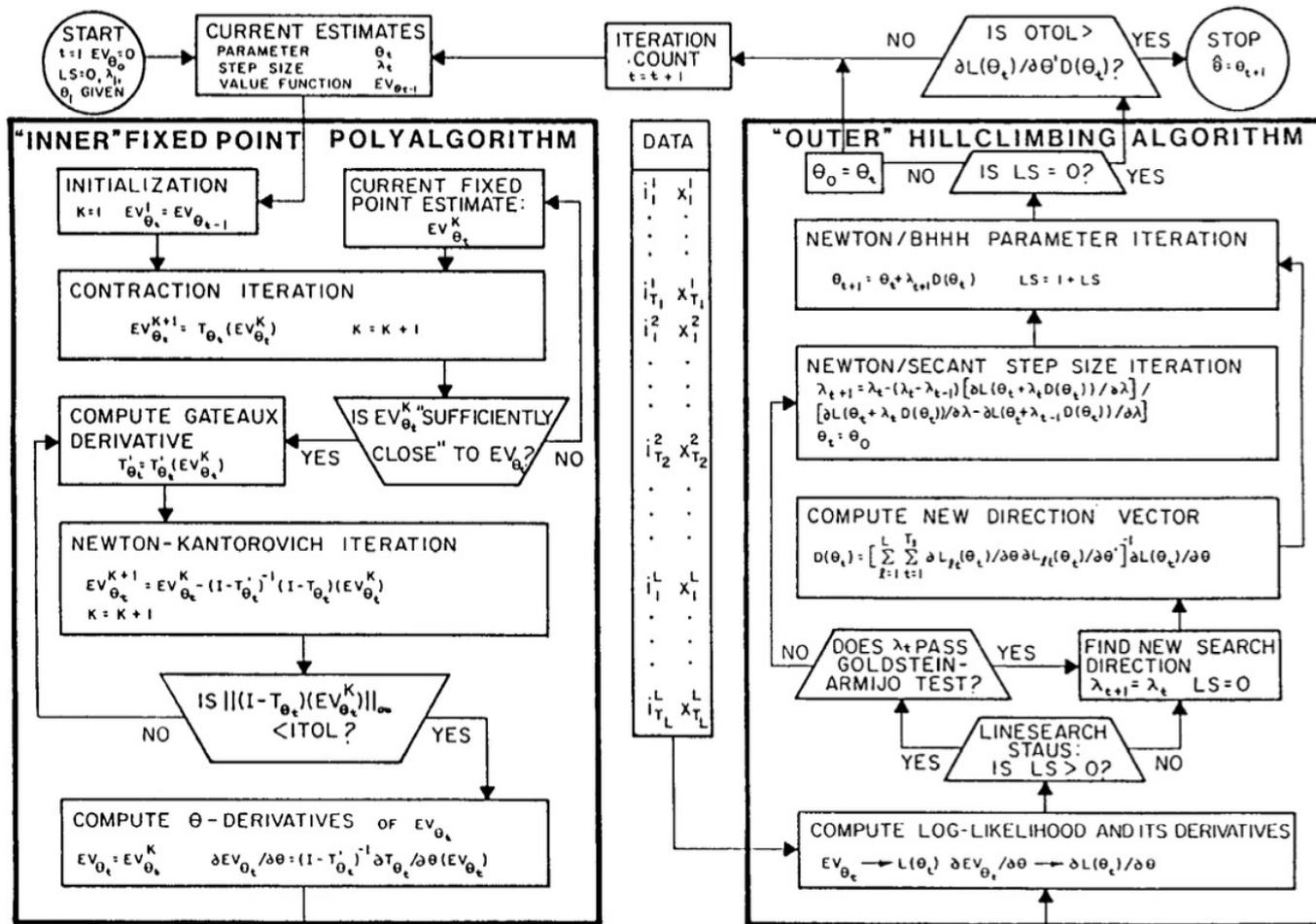


アルゴリズムの仕組み

(ネットワーク行動学 (羽藤) をもとに作成)

NFXPアルゴリズムの仕組み (詳細)

NESTED FIXED POINT MAXIMUM LIKELIHOOD ALGORITHM



不動点問題をとき、EVを求める

尤度関数の最大化を行う

不動点問題の解き方

$$T_{\theta}(EV_{\theta}) = \int_y \ln \left\{ \sum_{j \in C(y)} \exp\{u(x, j, \theta_1) + \beta EV_{\theta}(x, j)\} \right\} p(y|x, i, \theta_3) dy$$

↑上の式の不動点を求めたい.

関数 T_{θ} は縮小写像であることを利用する.

縮小写像(contraction mapping)とは

2点 x, y の間の距離を $d(x, y)$ で表す. このとき, 以下を満たす $0 \leq q < 1$ が存在するような写像を縮小写像という.

$$d(T(x), T(y)) \leq qd(x, y)$$

このことから, EV に何回も T を作用させることで, EV が収束し, 不動点が求まる.

バスのエンジン交換問題に適用する

紹介した解法を，先程のバスのエンジン交換問題に適用する．

まず，閉形式解が求められる手法が不適當であることを確認する

Table 3. Summary of replacement data (subsample of buses for which at least 1 replacement occurred)

bus group	交換までの走行距離 Mileage at Replacement				Elapsed time (months)				number of observations
	max	min	mean	standard deviation	max	min	mean	standard deviation	
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	273,400	124,800	199,733	37,459	74	38	59.1	10.9	27
4	387,300	121,300	257,336	65,477	116	28	73.7	23.3	33
5	322,500	118,000	245,291	60,258	127	31	85.4	29.7	11
6	237,200	82,400	150,786	61,007	127	49	74.7	35.2	7
7	331,800	121,000	208,963	48,981	104	41	68.3	16.9	27
8	297,500	132,000	186,700	43,956	104	36	58.4	22.2	19
full sample	387,300	82,400	216,354	60,475	127	28	68.1	22.4	124

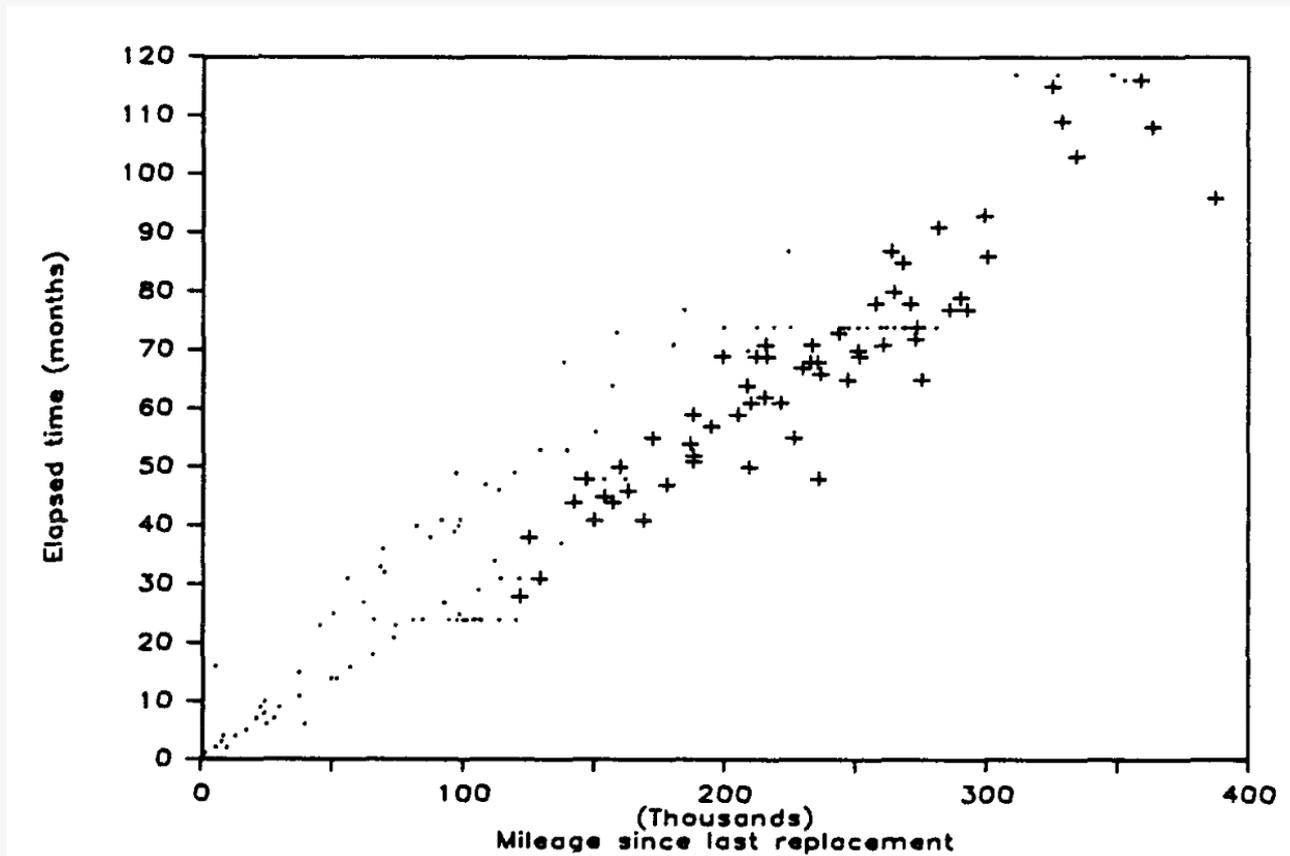
平均

標準偏差

バスのエンジン交換問題に適用する

紹介した解法を，先程のバスのエンジン交換問題に適用する．

まず，閉形式解が求められる手法が不適當であることを確認する



走行距離がある一定値を超えたら交換，というような運用はされていない

解くための準備

- 不動点計算をできるようにするため, x を離散化
- 関数形の指定を行う
 - $p(x_{t+1}|x_t, i, \theta_3)$ は多項分布とする.
 - $q(\epsilon_{t+1}|x_{t+1}, \theta_2)$ はガンベル分布とする
 - $c(x_t, \theta_1)$ は様々な関数形をおいて比較

結果

- $c(x_t, \theta_1)$ を様々な形にして尤度を求めた
 →尤度関数の大きさ, 他の非定量的な情報と矛盾しないかをもとに, 線形関数が適切と判断した.

Table 4. Summary of specification search*

Cost Function	Bus Group		
	1,2,3	4	1,2,3,4
cubic $c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2 + \theta_{13}x^3$	Model 1 -131.063 -131.177	Model 9 -162.988 -162.885	Model 17 -296.515 -296.411
quadratic $c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2$	Model 2 -131.325 -131.534	Model 10 -163.402 -163.771	Model 18 -297.939 -299.328
linear $c(x, \theta_1) = \theta_{11}x$	Model 3 -132.389 -134.747	Model 11 -163.584 -165.458	Model 19 -300.250 -306.641
square root $c(x, \theta_1) = \theta_{11}\sqrt{x}$	Model 4 -132.104 -133.472	Model 12 -163.395 -164.143	Model 20 -299.314 -302.703
power $c(x, \theta_1) = \theta_{11}x^{\theta_{12}}$	Model 5** N.C. N.C.	Model 13** N.C. N.C.	Model 21** N.C. N.C.
hyperbolic $c(x, \theta_1) = \theta_{11}/(91-x)$	Model 6 -133.408 -138.894	Model 14 -165.423 -174.023	Model 22 -305.605 -325.700
mixed $c(x, \theta_1) = \theta_{11}/(91-x) + \theta_{12}\sqrt{x}$	Model 7 -131.418 -131.612	Model -163.375 -164.048	Model 23 -298.866 -301.064
non-parametric $c(x, \theta_1)$ any function	Model 8 -110.832 -110.832	Model 16 -138.556 -138.556	Model 24 -261.641 -261.641

* First entry in each box is (partial) log likelihood value at $\beta = .9999$.
 Second entry is partial log likelihood value at $\beta = 0$.

** No convergence. Optimization algorithm attempted to drive $\theta_{12} \rightarrow 0$ and $\theta_{11} \rightarrow +\infty$.

結果

時間割引率

- $c(x_t, \theta_1)$ を線形とした上で、 β を変更したときの変化をみた。
- $\beta = 0$ のときと、 $\beta = 0.9999$ のときを比較すると、後者のほうがデータによくフィットした

Table 5. Structural estimates for cost function $c(x, \theta_1) = .001\theta_{11}x$. Fixed point dimension = 90 (standard errors in parentheses)

Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates/ Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic ($df = 4$)	Marginal Significance Level
$\beta = .9999$	<i>RC</i>	11.7270 (2.602)	10.0750 (1.582)	9.7558 (1.227)	85.46	1.2E-17
	θ_{11}	4.8259 (1.792)	2.2930 (0.639)	2.6275 (0.618)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3489 (.0052)		
	θ_{31}	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	<i>LL</i>	-2708.366	-3304.155	-6055.250		
$\beta = 0$	<i>RC</i>	8.2985 (1.0417)	7.6358 (0.7197)	7.3055 (0.5067)	89.73	1.5E-18
	θ_{11}	109.9031 (26.1630)	71.5133 (13.778)	70.2769 (10.750)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3488 (.0052)		
	θ_{31}	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	<i>LL</i>	-2710.746	-3306.028	-6061.641		
Myopia test:	LR Statistic ($df = 1$)	4.760	3.746	12.782		
$\beta = 0$ vs. $\beta = .9999$	Marginal Significance Level	0.0292	0.0529	0.0035		

政策分析への利用

- 政策変更の評価の1つとして、エンジン交換費用 $\bar{P} - \underline{P}$ を変化させることでバスのエンジン交換需要にどのような変化があるのかを分析する.
- $\bar{P} - \underline{P} = RC$ とおく.
- 1年間におけるエンジンの交換回数は、以下のように表される.

$$\bar{d}(RC) = \sum_{i=1}^{12} \sum_{m=1}^M i_t^m$$

- $\pi(x, i)$ が $\{i_t, x_t\}$ の定常分布だとすると、以下の式が成り立つ.

$$\pi(x, i) = \int_y \int_j P(i|x, \theta) p(x|y, j, \theta_3) \pi(dy, dj)$$

政策分析への利用

- 簡単のため、 $\bar{d}(RC)$ ではなく、その期待値である $d(RC) = E(\bar{d}(RC))$ を求めることを考える。
- 定常状態を考え、さらに $\{i_t^m, x_t^m\}$, $\{i_t^k, x_t^k\}$ が独立であると仮定すると、期待値は以下の式で表される。

$$d(RC) = 12M \int_0^{\infty} \pi_{\theta}(dx, 1)$$

(これを使ったときの買い替えまでの平均走行距離は、実測値とは大きく離れておらず、矛盾しない範囲である。)

政策分析への利用

エンジンコストは過去にあまり変化していないので、回帰曲線を書くのは難しい

バス1台のときの期待値を図示

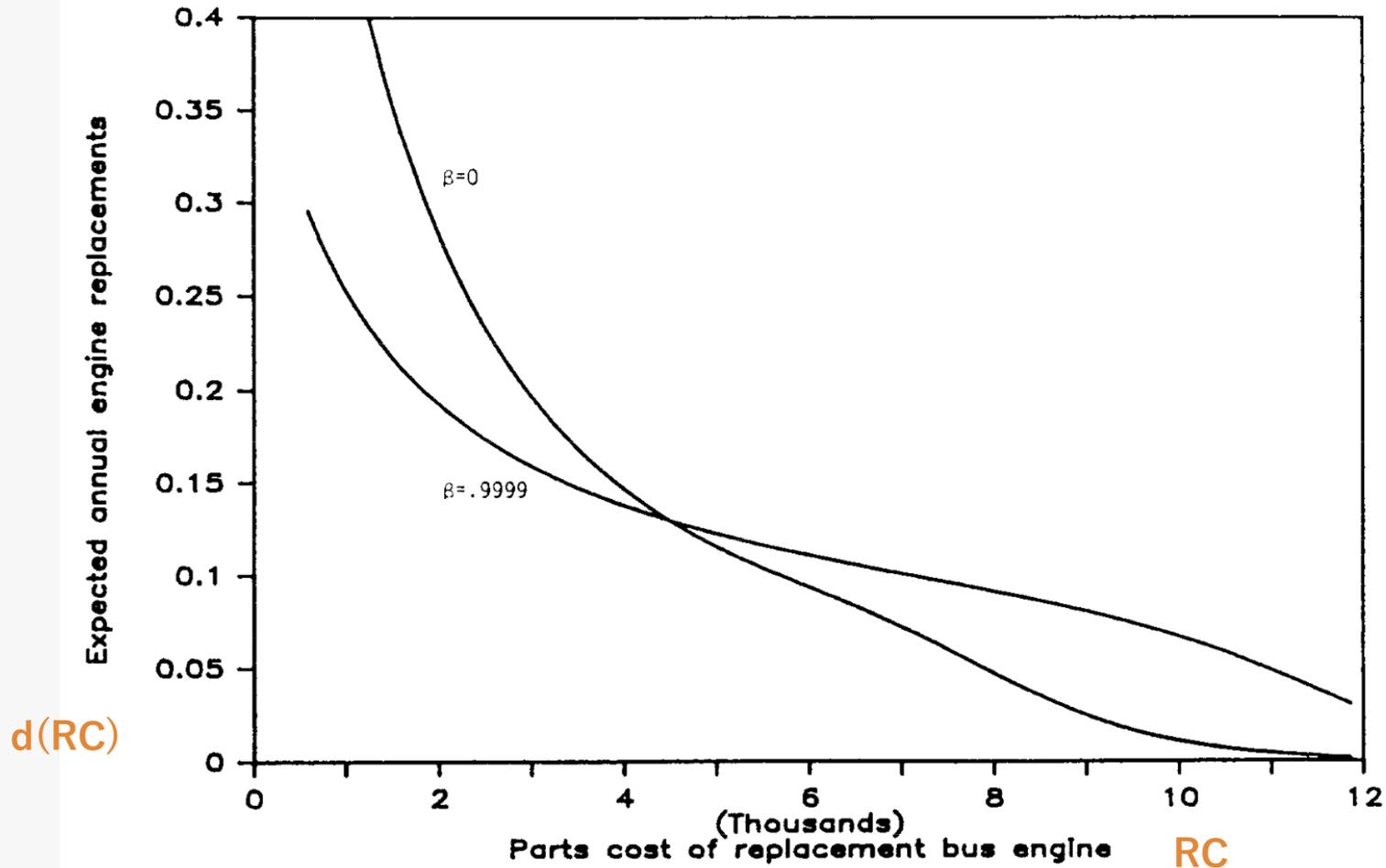


Fig. 7. Expected replacement demand function.

Conclusion

- 構造モデルは、データを制御された確率過程の実現であると解釈し、原理的なレベルから立式を行い、最適化問題を解くことで尤度関数を導出する
- 誘導モデルは、モデルの原理的なレベルのパラメータではなく、統合されたパラメータを用いて推定を行う。
- 構造モデルは構造の設定を誤ってしまうとうまく推定ができないため、過去の事象に対して誘導モデルを使うことで、誤った構造を仮定していないかを確認される。
- 構造モデルはパラメータがプリミティブなので、将来の政策の変更に対する影響を考えることができる
- 構造モデルは閉形式の解が得られないことが問題だったが、今回提案した手法により閉形式の解がなくとも推定ができるようになった。

参考文献

- Rust, John. "Statistical models of discrete choice processes." Transportation Research Part B: Methodological 22.2 (1988): 125-158.
- BinN Studies シリーズ「ネットワーク行動学」羽藤英二・伊藤創太・伊藤篤志 篇
(<http://bin.t.u-tokyo.ac.jp/kaken/data/full-20140926.pdf>)

発表で紹介した文献

- Berkovec J. and Rust J. (1985) A nested logit model of automobile holdings for one vehicle households. *TransDn.Res.* 19B. 275-285.
- Frank R. E. (1962) Brand choice as a probability process. *J. Business* 35, 43-56.
- Jeuland A., Bass F. and Wright G. (1980) A multibrand stochastic model compounding heterogeneous Erland timing and multinomial choice processes. *Operations Res.* B(2). 255-277.
- Manski C. F. and Sherman L. (1981) An empirical analysis of household choice among motor vehicles. *Trunspn.Res.* 14A, 349-366.
- Rust J. (1985a) Stationary equilibrium in a market for durable assets. *Econometrica* 53-54, 783-805.
- Rust J. (1985b) Equilibrium holdings distributions in durable asset markets. *Transpn. Res.* 19B. 331-345.
- Rust J. (1987a) Maximum likelihood estimation of discrete control processes. SSRI Working Paper 8107. University of Wisconsin, Madison, Wisconsin, *SIAM J. Control Optim.* (in press).
- Rust J. (1987d) Nested Fixed Point Algorithm Documentation. Wanaal. Software for IBM-PC, SSRI. Madison, Wisconsin.