異質性と動学化へのアプローチ

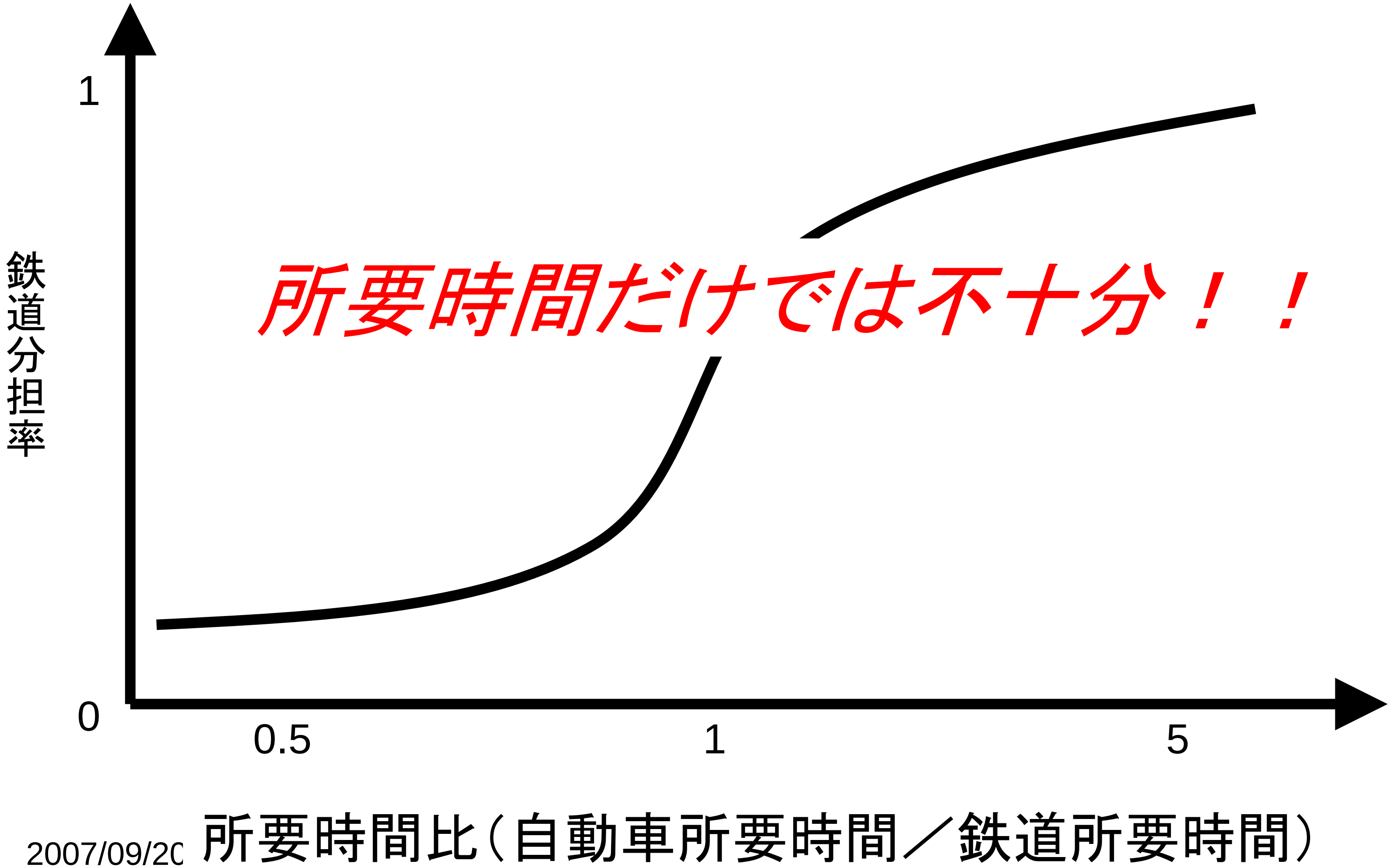
名古屋大学
山本俊行

内容

異質性の表現方法
- 選好の異質性
- 嗜好の異質性
- 意思決定方略の異質性

動学化の方法
- 変動について
- 複数時点での状態の記述
- 複数時点での変化の記述
- 連続時間軸上での変化の記述

四段階推計における分担率曲線



影響要因を追加する

- 費用を追加する
  - 所要時間の価値を算出する:労働に費やした場合に得られる収入を用いる
  - 所要時間と費用を合わせた一般化費用を分担率のX軸に用いる
- 運行頻度やアクセス距離などその他多くのサービス水準を追加する
  - 全ての要因を貨幣換算する？

最も基本的な離散選択モデル

- 様々なサービス水準の組み合わせからなる交通手段の効用を推定する（各サービス水準の重みを推定する）

$$U = a + b_1x_1 + b_2x_2 + \varepsilon$$

  - 線形効用関数を仮定する（補償型の効用）
  - 誤差項に確率分布を仮定する（分担率曲線と同様の曲線を生成）
  - 平均的な偏りは定数項で表す

地域によって定数項は異なる・・・異質性

定数項の構造化

- 地域・意思決定者による選択結果の違いを表現する

$$U = a + b_1x_1 + b_2x_2 + \underline{c_1z_1 + c_2z_2} + \varepsilon$$

  - 「定数項が属性によって異なる」と解釈できる
  - サービス水準の影響は属性によらず一定

時間価値は収入によって異なったりするはず

→ さらなる異質性の導入

重みの構造化

- サービス水準の重みを属性毎に異なるものとして推定する

$$U = a + f_1(z_1, z_2)x_1 + f_2(z_1, z_2)x_2 + c_1z_1 + c_2z_2 + \varepsilon$$

  - セグメント毎にサービス水準の重みを個別に推定する
  - 全てのサービス水準について異なる重みを仮定する場合は，セグメント毎に全く個別のモデルを推定することになる

はっきり属性は特定できないが，まだ重みの個別化が十分でない場合は？→→非観測異質性の導入

非観測異質性の導入1:潜在セグメントモデル

- 意思決定者が各セグメントに確率的に帰属することをモデル化

$$Pr(i) = \Sigma_j Q(j)P(i|j)$$

  - Q(j):セグメントjに帰属する確率(属性で構造化する場合と全員同じ確率を与える場合がある)
  - P(i|j):セグメントjに固有の効用関数によって選択肢 i を選択する確率

非観測異質性の導入2:
ランダム係数モデル

- サービス水準の重みを確率的に分布させる

$$U = a + f_1(z_1, z_2, \zeta_1)x_1 + f_2(z_1, z_2, \zeta_2)x_2 + c_1 z_1 + c_2 z_2 + \varepsilon$$

$$\Pr(i) = \int P(i|\, \zeta\,)g(\zeta)\partial\zeta$$

– g(ζ):仮定する確率分布(正規分布等)

最近では,確率分布の分散を属性で構造化するモデルも提案されている

2つの非観測異質性導入方法の相違

- 潜在セグメントモデル: 重みが離散分布
- ランダム係数モデル: 重みが連続分布

どちらも優劣はない

- どうせなら, 両方やってみますか？
- やってみた論文もあります. まぁモデルの適合度は上がってますけど･･･

ランダム係数モデルの若干の注意

- 正規分布を使うと，係数の正負は限定されない（定義域が無限なので必ず正負両方の値を取る）→時間とか費用とか普通は負であるはず
- 問題の生じない分布形が提案されている
  - 切断正規分布，対数正規分布，一様分布，三角分布，Johnson's S_B分布

$$c = a + (b - a) \cdot \frac{\exp(\xi)}{1 + \exp(\xi)}, \quad \xi \sim N(\mu, \sigma)$$

Mixed Multinomial Logit Modelの乱用

- ランダム係数モデルを含む Mixed multinomial logit model (MMNL) に対して，シミュレーションによる積分を活用した推定手法が急速に発展した
- そのため，近年多くの論文でMMNLモデルの適用が見られた

- 2006年IATBR会議では，非観測異質性で処理するのではなく，説明変数で構造化し，構造を理解しようという反省が出てきている

意思決定方略の異質性

- 線形効用関数は真の意思決定方略の近似
  - 例え真の意思決定方略が異なっていても，重み付き平均は最も選択の近似能力が高い
  - ただし，サービス水準間に負の相関がある場合には，再現力が低くなることが分かっている

真の意思決定方略をモデル化する価値があるかも

様々な意思決定方略

方略	概要
荷重加算型	各サービス水準の重み付き平均
等加算型	各サービス水準の単純平均
勝率最大化	各サービス水準毎に望ましい選択肢を選び，選ばれた回数が最も多い選択肢
連結型	全てのサービス水準が許容基準を満たす選択肢
分離型	サービス水準のうち一つでも希望水準を満たす選択肢
辞書編纂型	最も重要なサービス水準について最も望ましい選択肢
EBA	重要度によって確率的に選んだサービス水準が最低水準を満たす選択肢

選択の状況や意思決定者によって用いられる意思決定方略は異なる

複数の意思決定方略を含む潜在セグメントモデル

$$Pr(i) = \Sigma_j Q(j)P(i|j)$$

- P(i|j): セグメントjに固有の意思決定方略によって選択肢 i を選択する確率
- ただし，多くの意思決定方略は推定しづらい
- 可能性としてしては，
  - 選択結果以外の情報を取得し推定に用いる
  - ベイズ推定，データマイニング，等の推定手法の適用

選択結果以外の使える情報

- 使っている意思決定方略を直接アンケート調査等で聞く
- 意思決定の際に考慮した要因を聞く（非補償型の意思決定の多くは用いる情報を限定する）

選択結果とあわせてデータが得られる同時確率（尤度）を最大化する

ベイズ推定

$$P(\theta|X) = \frac{\Pr(X|\theta)\Omega(\theta)}{\int_{\theta=-\infty}^{\infty} \Pr(X|\theta)\Omega(\theta)d\theta}$$

概要

- パラメータの事前分布（通常は無知の分布）とデータが得られる尤度からパラメータの事後分布をベイズの定理によって求める
- 通常は解析的に解けないため，MMNLで用いられるのと同様のシミュレーションによって数値積分を行う

利点

- 一部のパラメータが推定不能の場合でも，他のパラメータを推定することが可能
- パラメータ空間が大局的に凸になっていない場合でも局所解に陥らない可能性が高い

データマイニング

概要

- 情報量基準等の指標と計算速度に優れたアルゴリズムを用いて，非線形関数やif-then ルールを効率的に探索する
- 様々なアルゴリズムによるツールがある

意思決定方略の探索に使えそうなツール

- CHAID, CART, C4.5等の決定木：確立されており利用が容易
- Bayesian Belief Network：より統計解析的

動学化へのアプローチ1:
変動について（北村, 2003）

交通は移ろいやすく, 交通現象に変動はつきもの

- 差異：一断面における個体間の変動
- 変化：特定の個体の系時的変動
- 変動：確率過程的変動

交通行動分析データの進展

- 断面データ
  - 環境の異なる個人間差異

↓

- パネルデータ
  - 時点間の個人の行動変化

↓

- 長期観測データ
  - 個人内変動の考慮

長期観測の利点

- 断面データ（1時点）

  $y = f(x) + e$(差異, 変化, 変動)

- パネルデータ

  $y_t = f(x_t) + e$(差異) $+ e_t$(変化, 変動)

  $y_{t'} = f(x_{t'}) + e$(差異) $+ e_{t'}$(変化, 変動)

  ⇩

  $y_{t'} - y_t = f(x_{t'}) - f(x_t) + e_{t'}$(変化, 変動) $- e_t$(変化, 変動)

長期観測の利点

- 長期観測

$y_{t1} = f(x_{t1}) + e$(差異) $+ e_t$(変化) $+ e_{t1}$(変動)

$y_{t2} = f(x_{t2}) + e$(差異) $+ e_t$(変化) $+ e_{t2}$(変動)

$y_{t'1} = f(x_{t'1}) + e$(差異) $+ e_{t'}$(変化) $+ e_{t'1}$(変動)

$y_{t'2} = f(x_{t'2}) + e$(差異) $+ e_{t'}$(変化) $+ e_{t'2}$(変動)

$E(y_{t'N}) - E(y_{tN}) = E(f(x_{t'})) – E(f(x_t))$

$+ e_{t'}$(変化) $- e_t$(変化)

長期観測データの活用

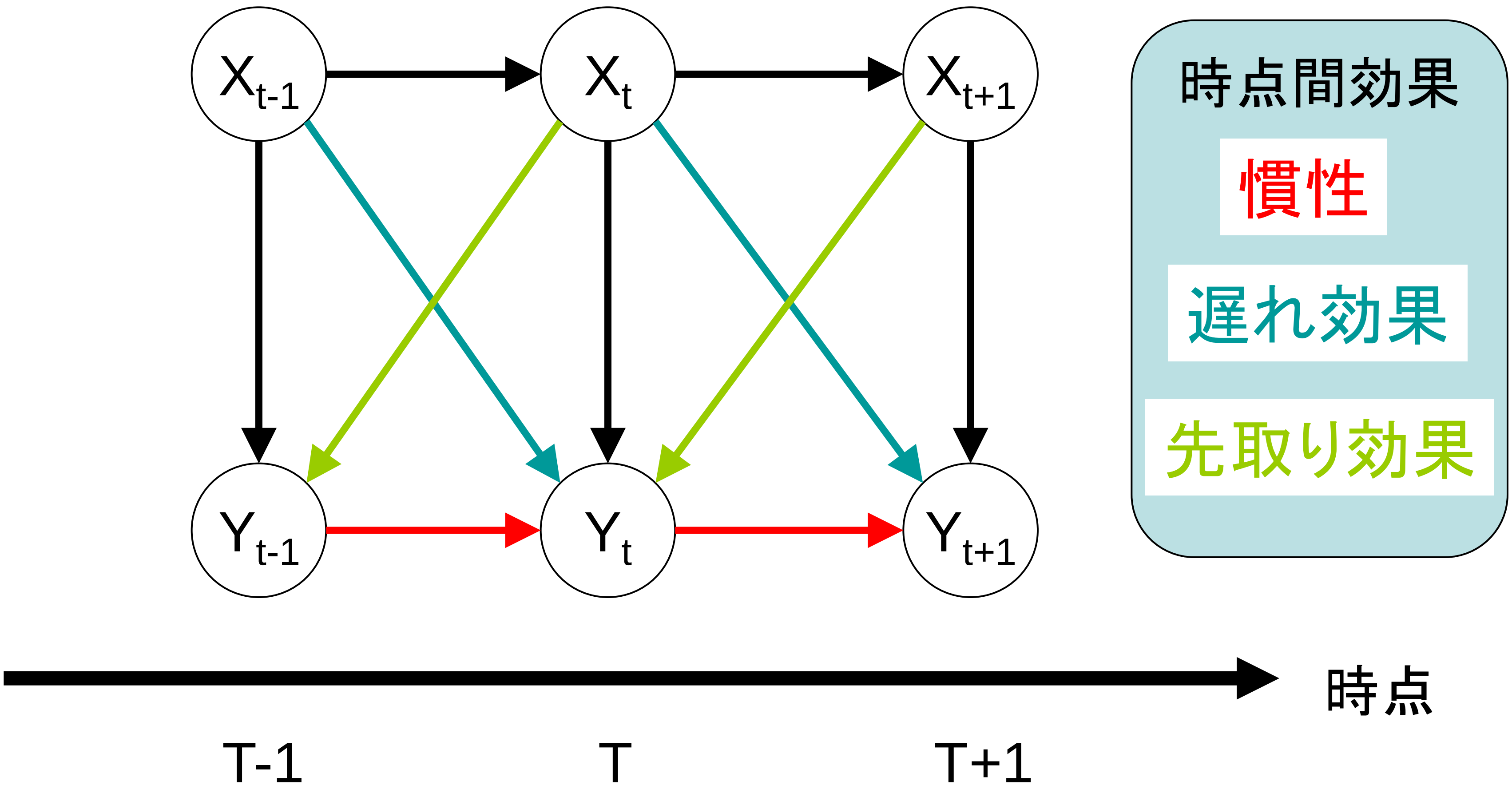
- 変動のモデル化
  - 変動を明示的に導入して日々の行動をモデル化
  - 行動の基にある，より安定的な意思決定原理のモデル化
- 変動の除去
  - 一定期間に集計した行動のモデル化
  - 一週間の時間配分モデル等への multiple discrete-continuous choice model や多変量頻度モデル等の適用

動学化へのアプローチ2

- 通常のモデルはある時点のサービス水準が同じ時点の交通手段選択を決定すると仮定している
- サービス水準が変化したら直ぐに行動も変化するのか？

動的モデル構築の必要性

複数時点での状態のモデル化

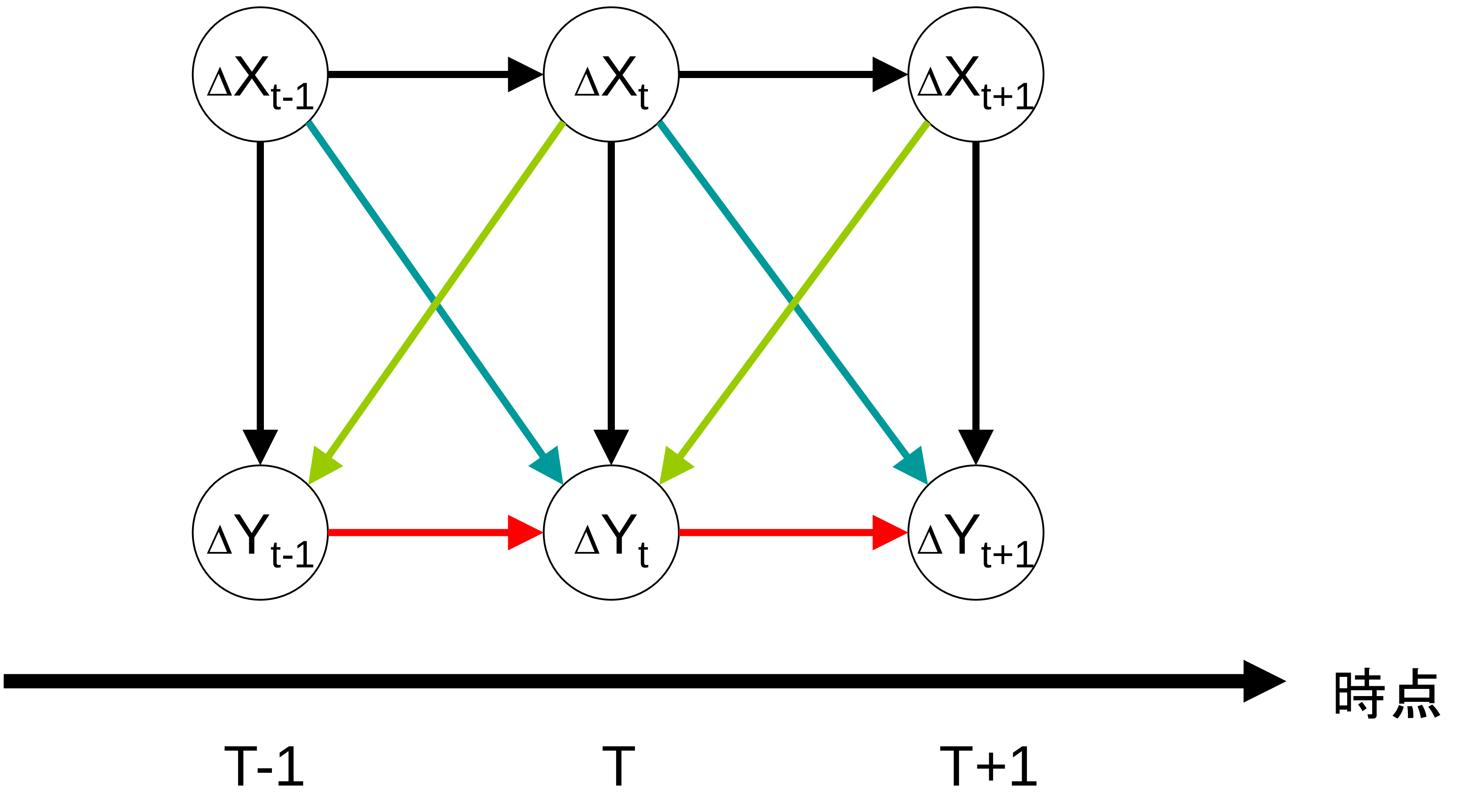


複数時点の状態モデルの問題点

$$Y_t = aY_{t-1} + bX_t + cX_{t-1} + dX_{t+1} + \varepsilon_t$$

- 誤差項は時点間で相関を持つ（非観測異質性は時点間で同一となる）
- 上記のようなモデル構造だと，X_tの値が増加した場合も減少した場合もY_tに与える影響の大きさは同じで正負対象（変化の対象性）
- 自動車保有台数等では，変化の対象性が成り立たないケースが多く見られる

複数時点での変化のモデル化



複数時点の変化モデルの問題点

$$\Delta Y_t = a\Delta Y_{t-1} + b\ \Delta X_t + c\ \Delta X_{t-1} + d\ \Delta X_{t+1} + \varepsilon_t$$

- 説明変数は ΔX_t の他に，X_t も含まれる
- 変化の非対称性も考慮可能

- 時間は連続なのに，離散的に扱っている
- 意思決定の時点とモデル化が整合していないケースも生じる

連続時間軸上での変化のモデル化

生存時間モデルの適用

- 連続時間を扱える
- 時点間の誤差項の相関等で苦労する必要がない
- 意思決定（選択）行動について，明示的には効用理論の枠組みで表現していない
  – 効用理論との整合性について示した論文もあり関連付けることは可能（小林ら, 1997; 佐々木ら, 1997）

生存時間モデル

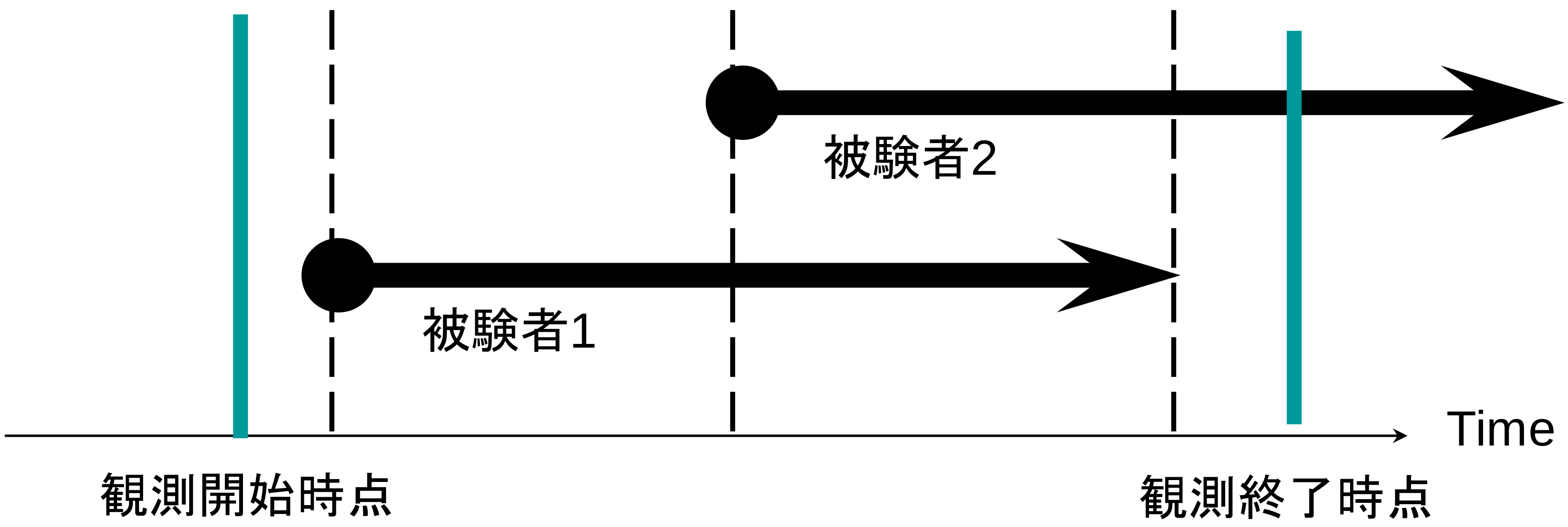
ハザード関数：t 以前に事象が発生していないという条件下で t に事象が発生する条件付き確率密度

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\left(t + \Delta t > T \geq t \middle| T \geq t\right)}{\Delta t} = \frac{f(t)}{1 - F(t)}$$

$f(t)$:　確率密度関数
$F(t)$: 累積分布関数

観測打ち切り

- 事象が最後まで発生しないケースも解析可能
  – データから削除すると推定パラメータのバイアスにつながる



$$S(t) = \Pr(T \geq t) = 1 - F(t)$$

尤度関数

$$L = \prod_i \left\{ f(t)^{(1-\delta_i)} S(t)^{\delta_i} \right\}$$

$$= \prod_i \left\{ h(t)^{(1-\delta_i)} S(t) \right\}$$

δ_i: 事象が観測できなかったことを表すダミー変数

以下のように, $S(t)$ が $h(t)$ で表せることから上記は $h(t)$ の関数

$$S(t) = \exp\left(- \int_{s=0}^{t} h(s) ds \right)$$

ハザード関数の構造化

- 比例ハザードモデル：基準ハザード関数に異質性が影響を与える．時点間で一定

$$h(t) = h_0(t) r(\theta X)$$

- 加速故障モデル：時間の進み方に異質性が影響を与える

$$h(t) = h_0\{r(\theta X) t\}$$

構造化の方法によるモデルの分類

モデル	基準ハザード	異質性
ノンパラメトリックモデル	ノンパラメトリック	ノンパラメトリック
セミパラメトリックモデル	ノンパラメトリック	パラメトリック
パラメトリックモデル	パラメトリック	パラメトリック

観測精度との整合性

- 通常の調査では，変化の時点を連続的に観測することは稀であり，通常は1年に1度等のパネル調査が用いられる
- 生存時間モデルでは「変化」がある時間内に生起したことを表現することも生起確率を t について積分することで表現可能（積分区間は任意に設定可能）

$$\Pr\left(t_1 < s \le t_2\right) = \int_{s=t_1}^{t_2} f\left(s\right) ds$$
$$= F\left(t_2\right) - F\left(t_1\right)$$
$$= S\left(t_1\right) - S\left(t_2\right)$$

おわりに

- 時間をかければモデルの適合度は向上する
  - このプレゼンファイルも火曜日より枚数が増えることは増えた
- モデルを分かりやすくするために，本質的なこと以外は削除することも大事
  - このプレゼンファイルの内容は枚数を増やしたことで分かりにくくなったかも知れない
- ある程度の時間であきらめる勇気も必要：あきらめないと2008年度モデルの新車は出ない
  - もちろん言い訳にしてはいけませんが・・・