

# イベント・ヒストリー・アナリシスの成行き調査分析への応用

—— 比例ハザードモデルを用いて ——

津富 宏

成行き調査は、私たちの矯正処遇を客観的に評価するための貴重なデータを与えてくれる。イベント・ヒストリー・アナリシスは、時間の流れ（ヒストリー）の中で生じるできごと（イベント）の生起する過程を、確率過程としてモデリングし、また、イベントの生起に關与する要因をとらえるための統計手法（Allison, 1984；Tuma and Hannan, 1984；Kalbfleish and Prentice, 1980；Cox and Oakes, 1984）である。本稿では、イベント・ヒストリー・アナリシスの一手法である比例ハザードモデルにより、成行き調査を分析し、処遇効果を評価する方法を紹介する。

## イベント・ヒストリー・アナリシスの基本的構造

成行き調査で得られる情報の中でも、特に、矯正職員の関心を惹くイベントは、矯正施設への再入であろう。ある刑務所から、出所した受刑者について、その成行きを追跡したとしよう。時点  $t$  までに再入していなかった受刑者が、時点  $t$  に再入する瞬間的確率（instantaneous rate）は、ハザード・レイト（ $h(t)$ ）と呼ばれ、イベント・ヒストリー・アナリシスでは、ハザード・レイトをモデリングすることにより、受刑者が再入に至る過程が表現される。ハザード・レイトは、イベントの生起する時点をランダムな変数  $T$  で表わすと、次のように与えられる。

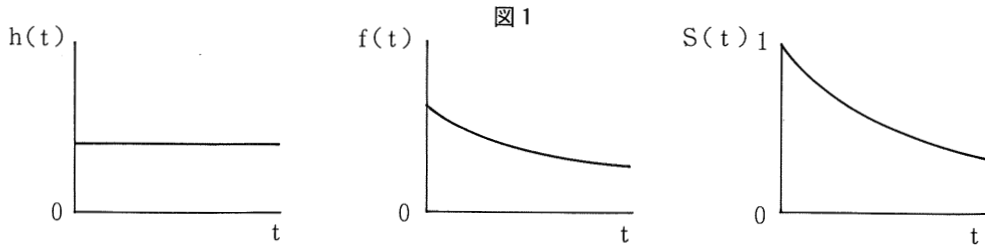
$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t > T \geq t \mid T \geq t)}{\Delta t} = f(t)/S(t) \quad (1)$$

$f(t)$  は、受刑者が、時点  $t$  までに、再入していたかどうかにかかわらず、時点  $t$  に再入する瞬間的確率を表わす確率密度関数（probability density function）、 $s(t)$  は、受刑者が時点  $t$  まで再入しない確率を表わす生存関数（survivor function）であり、それぞれ、次のように与えられる。

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t > T \geq t \mid T \geq t)}{\Delta t} = -d[S(t)]/dt = h(t) \exp\left[-\int_0^t h(u) du\right] \quad (2)$$

$$S(t) = P(T \geq t) = 1 - \int_0^t f(t) dt = \exp\left[-\int_0^t h(u) du\right] \quad (3)$$

(2), (3)式に示されているように、 $f(t)$ ,  $s(t)$  は、それぞれ、 $h(t)$  から決定される。たとえば、 $h(t)$  が時間にかかわらず一定であれば、 $f(t)$ ,  $s(t)$  は、図1のような変化を示すであろう。



$h(t)$ が一定であるということは、すべての時点において、その時点までに再入していない受刑者の再入する可能性が等しいということである。話を単純化するために、時間を非連続と見なした場合、 $h(t)$ が一定であれば、受刑者100人のうち、再入していない者の数が、100人、90人、81人、72.9人と減少していくことになる。時間を連続と見なした場合は、図1に示される通りである。

### 比例ハザードモデル

イベント・ヒストリー・アナリシスにおける代表的なモデルは、比例ハザードモデルと accelerated failure-time model である (Kalbfleish and Prentice, 1980)。説明変数のハザード・レイトに対する効果の分析には、比例ハザードモデルが、時間のハザード・レイトに対する効果の分析には、accelerated failure-time model が適しており (Yamaguchi, 1987)、矯正処遇は、再入にかかわるハザード・レイトの説明変数であるから、処遇効果の分析には、比例ハザードモデルが適切である。

比例ハザードモデルでは、ハザード・レイトは、共変数 (説明変数)  $x$  の対数線形関数として、次のようにモデリングされる。

$$h(t) = r(t) \exp\left[b_0 + \sum_{j=1}^J b_j x_{ij}(t)\right] = r(t) \exp\left[b_0 + b' X_i(t)\right] \quad (4)$$

あるいは、

$$\log h(t) = \log r(t) + b_0 + \sum_{j=1}^J b_j x_{ij}(t) = \log r(t) + b_0 + b' X_i(t)$$

$r(t)$ は、baseline hazard function といわれ、時点  $t$  の関数で、ハザード・レイトが、時間に依存する関数であることを示している。また、 $x_{ij}(t)$ は、時点  $t$  における、個人  $i$  にかかわる、 $J$ 個の共変数 (covariate) ( $X_i(t)$ は、 $x_{ij}(t)$ を要素とするベクトル) で、ハザード・レイトが個々人の特性に依存することを示している。(4)式に示されるように、共変数は、(性別のように) 時間によって変化しなくても (time-invariant)、(薬物使用のように) 時間によって変化しても (time-varying) よい。特定の矯正処遇をうけたかいないかは、施設から釈放される時点では既に確定しているから、再入に関する比例ハザードモデルでは、矯正処遇は、時間によって変化しない共変数である (これに対し、釈放後の就業状態は、時間によって変化する共変数である)。

たとえば、 $x_{i1}$ を、受刑者  $i$  が、特定の処遇をうけたかいないかを表わす共変数としよう。処遇

の有無  $x_{i1}$  が、ハザード・レイトに影響を与えているという比例ハザードモデルは、次のようになる。

$$\log h(t) = \log r(t) + b_0 + b_1 x_{i1}$$

このモデルに、処遇以外で、再入に影響を与えるであろう要因、たとえば、年齢・暴力団加入歴の有無・薬物使用歴の有無などを、共変数として含めることで、成行き調査から得られるデータを、処遇効果分析に用いることが可能となる。

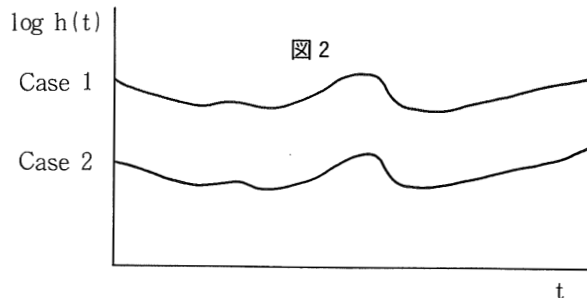
すべての共変数が、時間によって変化しない場合には、どの時点においても、個人間のハザードの比は一定であり、このため、このモデルは、比例ハザードモデルの名がある。すなわち、ある2人の受刑者のハザードの比は、どの時点においても、

$$h(t)_1/h(t)_2 = \frac{r(t)\exp[b_0 + b'X_1]}{r(t)\exp[b_0 + b'X_2]} = \frac{\exp[b'X_1]}{\exp[b'X_2]} = c \quad (5)$$

あるいは、

$$\begin{aligned} \log h(t)_1 - \log h(t)_2 &= (\log r(t) + b_0 + b'X_1) - (\log r(t) + b_0 + b'X_2) \\ &= b'X_1 - b'X_2 \\ &= \log c \end{aligned}$$

と、一定であり、 $\log h(t)_1$ と $\log h(t)_2$ を時間の関数として、プロットすると、図2のようになる。この比例ハザードの前提は、どの時点においても、共変数のハザード・レイトに対する効果が変わらない、すなわち、共変数のハザード・レイトに対する効果と時間のハザード・レイトに対する効果が、交互作用をもたないという前提である。



もちろん、現実のデータに関して、この比例ハザードの前提が成り立つとは限らない。たとえば、矯正処遇が、再入についてのハザード・レイトに対して与える効果が、釈放後経過した時間によって異なるというのは、非現実的である。ある処遇（たとえば、職業訓練）を受けた受刑者のハザード・レイトと、その処遇を受けなかった受刑者のハザード・レイトの比が、釈放直後に、1：3であったとしても、5年後には、1：1となるかもしれない。このような場合には、時間と共変数の関数、 $g(x, t)$ を想定し、これを時間によって変化する変数としてモデルに取り込むことで、比例ハザードの前提を緩和したモデリングが可能である（Allison, 1984）。

### r(t)のモデリングに伴う諸問題—準パラメトリックなモデルの有効性—

r(t)は、tの関数として、さまざまにモデリングが可能である。たとえば、 $r(t) = \alpha$ とすれば指数モデル、 $r(t) = \alpha t^{(\beta-1)}$ とすればワイブル (Weibull) モデル、 $r(t) = \alpha * \exp [(\beta - 1)t]$ とすればゴンペルツ (Gompertz) モデルなどの、さまざまなモデルが想定できる。しかしながら、残念なことに、r(t)について、どのモデルを選択すべきかについての理論的根拠はなく、特に、ハザード・レイトが単調に増加する場合及び減少する場合を除いては、適当なr(t)が存在しないこともある。しかも、誤ったモデルを選択した場合には、共変数のハザード・レイトに対する効果bの推定値が、バイアスされることが知られている (Tuma and Hannan, 1984)。

さらに、r(t)を特定化すると、サンプルが均質でなく、観測されない不均質さ (unobserved heterogeneity) をもっている場合には、bの推定値がバイアスされることが知られている。観測されない不均質さを、誤差項としてモデリングする試み (e.g. Trussell and Richards, 1985) が行われてきたが、いまだに適切な方法は見いだされていない (Palloni and Sørensen, in press)。

幸いなことに、比例ハザードモデルは、r(t)を特定化せずにモデリングし、bの推定値を得ることが可能である。このr(t)を特定化しないモデリングは、準パラメトリック (semi-parametric) なモデリングといわれるが、処遇効果研究のように、時間のハザード・レイトに対する効果ではなく、共変数 (すなわち、処遇の有無) のハザード・レイトに対する効果を問題とする研究においては、r(t)を特定化せずに、準パラメトリックなモデルを推定することが望ましい。<sup>1)</sup> 準パラメトリックなモデルの推定法は、Cox (1975) の偏尤法として知られている。

#### Coxの偏尤法

r(t)を規定するパラメータ値を推定するには、イベントの生起した時点が必要だが、共変数の効果bを推定するには、イベントの生起した順序があればよい。Coxの偏尤法は、この点に着目した、イベントの生起した順序をデータとして用いる、準パラメトリックなモデルの推定法である。偏尤法は、bのかかわる部分の尤度、すなわち、偏尤度 (partial likelihood) に着目することにより、漸近的に (サンプルが大きくなるにつれ) バイアスされず、かつ、漸近的に効率的なbの推定値を与えてくれる。<sup>2)</sup>

偏尤法では、偏尤度を最大にするbの値が求める推定値である。偏尤度 PLは、それぞれのイベントが観測される確率を、すべて掛け合わせたもので、観測されたイベントの回数をKとすると、

$$PL = L_1 * L_2 * \dots * L_K = \prod_{k=1}^K L_k \quad (6)$$

である。受刑者5人 (それぞれ、 $p_1, p_2, p_3, p_4, p_5$ とする) の成行きを2年間追跡したところ、 $p_2, p_5, p_3$ の順で再入し、他の2人は再入しなかったとしよう。一番目のイベント (以下、イベント1) が時点 $t$ におこったとしよう。イベント1に関する偏尤度 $L_1$ は、5人すべてが再入するリスクがある場合に、(他の4人ではなく)  $p_2$ が再入する確率として定義される。すなわち、

$$L_1 = \frac{h_2(\hat{t})}{h_1(\hat{t}) + h_2(\hat{t}) + h_3(\hat{t}) + h_4(\hat{t}) + h_5(\hat{t})} \quad (7)$$

$r(t)$ は、どの受刑者についても一定だから、約分され、

$$L_1 = \frac{\exp[b' X_2]}{\exp[b' X_1] + \exp[b' X_2] + \exp[b' X_3] + \exp[b' X_4] + \exp[b' X_5]}$$

同様に、イベント2，イベント3に関する偏尤度は、

$$L_2 = \frac{\exp[b' X_3]}{\exp[b' X_1] + \exp[b' X_3] + \exp[b' X_4] + \exp[b' X_5]} \quad (8)$$

$$L_3 = \frac{\exp[b' X_4]}{\exp[b' X_1] + \exp[b' X_3] + \exp[b' X_4]} \quad (9)$$

で与えられる。よって、

$$PL = L_1 * L_2 * L_3 \quad (10)$$

となり、PLを最大化するようなbの値が求める推定値である。イベント・ヒストリー・アナリシスの特徴は、観察期間（本例の場合、2年間の追跡期間）の終了により観察の打ち切られたケース（censored case）に関する情報をも、分析に用いることができる点にあり、偏尤法では、分母にこれらのケースに関する情報が盛りこまれている。

### データへの適用

以上のモデルを、実際にデータに適用してみよう。サンプルは、ある少年院の仮退院者217名から、他施設からの保護移送者、成行きの不明な者、収容継続を言い渡され20才を過ぎて出院した者を除いた206名で、彼らが、20才になるまでの間、出院から何日たって、矯正施設への再収容の言い渡しというイベントを経験したかが関心の対象である。<sup>3)</sup> 共変数は、処遇を代表するものとして中間期にどの寮に属したか（ryou），非行性を代表するものとして施設歴があるかないか（sisetsu）を用いた。すなわち、次のモデルが推定された。

$$\log h(t) = \log r(t) + b_0 + b_1 * ryou + b_2 * sisetsu$$

モデルの推定は、統計パッケージLIMDEPを用いることにより、次のように行われた。<sup>4)</sup> LIMDEPを用いて、イベント・ヒストリー・アナリシスを行うには、次のような形のデータが必要である。仮に、4人分のデータを示そう。

| time | status | ryou | sisetsu |
|------|--------|------|---------|
| 183  | 0      | 1    | 0       |
| 285  | 1      | 0    | 0       |
| 307  | 1      | 0    | 0       |
| 252  | 0      | 0    | 1       |

time は、 入院日から観察の打ち切られた日までの日数を示している。これは、再収容の言い渡しを受けた少年については、 入院から再収容の言い渡しを受けるまでの日数であり、 言い渡しを受けなかった少年については、 入院から 20 才の誕生日の前日（あるいは死亡した日）までの日数である。status は、 再収容の言い渡しを受けた少年については 1、 その他の少年については 0。ryou は、 ある寮に属していれば 1、 他の寮であれば 0。sisetsu は、 教護院歴・少年院歴があれば 1、 なければ 0 である。

データファイル名を、 NARIYUKI. DAT とすると、

```
READ; FILE=NARIYUKI. DAT; NOBS=206; NVAR=4; NAMES=1 $
```

というコマンドで、 データが読み込まれる。NOBS はケースの数、 NVAR は変数の数、 NAMES はデータファイルにおいて変数名の占める行数である。さて、 Cox の偏尤法による準パラメトリックなモデルの推定は、 次のコマンドで行われる。

```
SURVIVAL; LHS=time,status; RHS=ryou,sisetsu $
```

LHS は left-hand side、 RHS は right-hand side の略であり、 共変数は RHS に列挙される。推定結果は、 表 1 に示すとおりである。<sup>5)</sup>

|         | 係数    | T値     |
|---------|-------|--------|
| ryou    | -.625 | -1.528 |
| sisetsu | .553  | 1.565  |

ryou の効果は、 通常用いられる 5%水準では有意ではないが、 評価研究の最初のステップとして、「有望な」処遇を発見するという目的でみれば、 13%を上回る水準で有意であり、 処遇改善の端緒が見いだされたと考えられる。 $\exp[-.625] = .5353$  であり、 この寮で処遇を受けた少年は、 他の寮で処遇を受けた少年に比べ、 ハザード・レイト、 すなわち、 再入する瞬間的確率が  $(100-54=)$  46%も低いことがわかる。

以上、 イベント・ヒストリー・アナリシスを用いて、 成行き調査を分析し、 処遇効果を検討する方法を示したが、 本分析は、 実例を、 統計上わかりやすい形で示すためのもので、 上記の結果を、 実際の処遇と結びつけて解釈することは適当ではない。実際の処遇評価研究にあたっては、 さらに多くのコントロール変数を導入することが必要である。

## おわりに

従来、成行き調査から得られるデータの分析は、再入の有無を決定する要因の分析に過ぎなかったが、<sup>6)</sup>再入というイベントの生起を時間の流れの中でモデリングするイベント・ヒストリー・アナリシスは、再入の順序や時期を決定する要因の分析を可能とする。

この意味で、イベント・ヒストリー・アナリシスは、成行き調査から得られる情報を最大限に活用した統計手法であり、その利用により、従来の手法では見いだしえなかった、処遇効果を見いだしていくことも可能である。筆者は、イベント・ヒストリー・アナリシスが、矯正職員によって、広く利用され、成行き調査分析の標準的手法として定着していくことを期待している。

中でも、比例ハザードモデルによるイベント・ヒストリー・アナリシスは、処遇効果の分析に適し、コンピュータパッケージの入手可能性の点でも優れており（IBM 互換機上で動作するイベント・ヒストリー・アナリシスのパッケージのレビューは、Goldstein et al., 1989 をみよ）、矯正職員がまず理解すべき手法である。

また、イベント・ヒストリー・アナリシスの標準的手法ではないため、本稿ではとりあげなかったが、時間を連続した（continuous）ものではなく、非連続な（discrete）ものとして扱うディスクリット・メソッド（Allison, 1982）も、構造の簡明さ及びパッケージの入手可能性の点で優れており、矯正現場におけるデータ分析に適切であろう。

イベント・ヒストリー・アナリシスは、イベントが2種以上ある場合（たとえば、保護観察と少年院送致）の分析、繰り返しおこりうる場合（たとえば、逮捕）の分析などに拡張でき、成行き調査における処遇効果の検討ばかりでなく、犯罪・非行行動の研究に豊かな可能性をもっている。たとえば、収容者の生育歴、職業歴、学歴、薬物使用歴、反社会集団加入歴などのイベント・ヒストリーを分析の対象とすることにより、犯罪・非行行動に関する新しい知見を見だし、それを矯正処遇の改善に役立てることも意義深い。イベント・ヒストリー・アナリシスは、新しい統計手法であり、未完成の部分も多い。統計手法自体の発展に目を配りつつ、よりよき応用を目指していくことが大切である。

## 謝 辞

本稿の作成にあたっては、御協力いただいた施設の方々の御援助により、貴重なデータを利用させて頂くことができました。ここに謹んで、御礼申し上げます。

## 注

1) 準パラメトリックなモデルに対し、（比例ハザードモデルに限らず）その他のイベント・ヒストリー・アナリシスのモデルは、モデル全体をパラメータ化するという意味で、全パラメトリックな（fully-parametric）モデルといわれる（全パラメトリックなモデルの犯罪研究への応用は、原田（1989）で紹介されている）。なお、処遇効果分析ではなく、成行き予測を行う場合には、全パラメトリックなモデルの方が、準パラメトリックなモデルよりも、ストレートに適用できるが、準パラメトリックなモデルを用いて予測を行うことも可能である（e. g. Breslow, 1974）。

2) Cox の偏尤法は、パーソナルコンピュータの処理能力の範囲で、時間によって変化する共変数を取り込んだモデルを、容易に推定できる点でも優れている。実際、入手可能な統計パッケージのほとんどは、偏尤法を用いた場合にのみ、時間によって変化する変数を許している（see Goldstein et al., 1989）。イベント・ヒストリー・アナリシスの長所は、時間によって変化する共変数をモデリングできる点にあり（Allison, 1984; Yamaguchi, 1987）、偏尤法は、この点でも、イベント・ヒストリー・アナリシスの長所を引き出す推定法である。たとえば、

Allison (1984) は、偏尤法を用いて、就業の逮捕に対する影響を分析し、1年間に何週間働いたかという、時間によって変化しない共変数を用いた場合には、就業が逮捕に影響を与えないにもかかわらず、それぞれの週に働いたかいないかという、時間によって変化する共変数を用いた場合には、就業が逮捕を減少させていることを見だしている。

3) もちろん、犯罪を行ったからといって、刑事司法機関の目に止まるとは限らず、まして、矯正施設への再入に至るとは限らない。Holden (1985) は、個々の犯罪が発覚する確率が小さければ、初めての逮捕をイベントとする比例ハザードモデルは、(警察による取り扱いにバイアスがない限り)一定の時間内に行われる犯罪の回数を説明するモデルとして解釈できることを示しており、同様に、初回再入に関する比例ハザードモデルは、犯罪の回数に関するモデルとして解釈できるであろう。

4) LIMDEP は、計量経済学の代表的統計パッケージの一つであり、IMB 互換のパソコン上で動作する。回帰分析を初めとして、choice-based sampling, probit, logit, ポワソン回帰分析, tobit, サンプルセレクションモデルなどの犯罪学研究にも利用度の高いさまざまな統計モデルが推定でき、かつ、同種のパッケージと比べるかに安価(\$350)である。入手先は、

Econometric Software, Inc.  
43 Maple Ave.  
Bellport, NY 11713

5) 比例ハザードの前提が守られているかどうか確認するため、ryou と時間の交互作用項を時間によって変化する共変数として含めた次のモデルを推計した。

SURVIVAL ; LHS=time,status ; RHS=ryou,sisetsu ; TVC=ryou\*(time)\$

このモデルを推定した結果、時間によって変化する共変数の効果は有意でなく、本稿では、表1の結果のみを報告した。

6) 再入の有無しか情報としてえられない場合の分析には、日本の矯正処遇評価研究には定着していないが、クロス表分析の手法である対数線形分析(津富, 1989), あるいは、従属変数がタミー変数である場合の回帰分析である logit analysis 及び probit analysis (Aldrich and Nelson, 1984) が適切である。

## 引用文献

- Aldrich, J. H., and F. D. Nelson. 1984. Linear Probability, Logit and Probit Models: Quantitative Applications in Social Sciences. Beverly Hills, Calif.:Sage.
- Allison, P. 1982. "Discrete-Time Methods for the Analysis of Event Histories." Sociological Methodology 1982. San Francisco: Jossey-Bass.
- , 1984. Event-History Analysis: Quantitative Applications in Social Sciences. Beverly Hills, Calif.:Sage.
- Breslow, N. E. 1974. "Covariance Analysis of Censored Survival Data." Biometrics 30:89-100.
- Cox, D. R. 1975. "Partial Likelihood." Biometrika 62:269-76.
- , and D. Oakes. 1984. Analysis of Survival Data. New York:Chapman and Hall.
- Goldstein, R., J. Anderson, A. Ash, B. Craig, D. Harrington, and M. Pagano. 1989. "Survival Analysis Software on MS/PC-DOS Computers." Journal of Applied Econometrics 4:393-414.
- 原田豊. 1989, "非行経歴研究へのイベント・ヒストリー・アナリシスの適用に関する諸問題." 科学警察研究所報告 30(1):57-68.
- Holden, R. T. 1985. "Failure Time Models for Thinned Crime Commission Data." Sociological Methods and Research 14:3-30.
- Kalbfleish, J. D. and R. L. Prentice. 1980. The Statistical Analysis of Failure Time Data. New York:John Wiley.
- Palloni, A., and Aage C. Sørensen. In press. "Methods for the Analysis of Event History Data: A Didactic Overview." In Baltes, P. B. et al. (eds.), Life-Span Development and Behavior (Vol. 10). Hillsdale,



- N.J.:Lawrence Erlbaum Associates.
- Trussell, J. T. and T. Richards. 1985. "Correcting for Unmeasured Heterogeneity in Hazard Models Using Heckman-Singer Procedure." Sociological Methodology 1985. San Francisco:Jossey-Bass.
- 津富宏. "対数線形モデルの矯正処遇効果分析への応用." 矯正教育研究 35 : 187-198.
- Tuma, N. B., and M. T. Hannan. 1984. Social Dynamics. New York:Academic Press.
- Yamaguchi, K. 1987. "Event-History Analysis:Its Contributions to Modeling and Causal Inference." 理論と方法 2:61-82.