

## 〔研究ノート〕

# ライフイベントの統計分析

## —イベントヒストリー分析—

中井 美樹\*

計量社会学や計量的・統計的手法への関心はこれまで社会科学の領域では決して少なくはなかった。しかし、「統計」的な方法に対するとっつきにくさや誤解が、こうした手法を研究に適用するうえでの妨げになってきた可能性がある。本稿では、計量社会学において幅広く用いられるイベントヒストリー分析をとりあげ、その方法について紹介する。イベントヒストリー分析の手法の特徴、考え方だけでなく、この手法を用いた先行研究にも言及しつつ、そこでの応用においてこの手法が社会学的研究に対してもつ利点を解説する。

キーワード：イベントヒストリー分析、ライフイベント、生存関数、センサリング（中途打ち切り）データ、結婚年齢

### 1. イベントヒストリー分析とは

イベントヒストリー分析とはどのような分析なのだろうか。まさにその名のとおり、「イベント」すなわち“出来事”の「ヒストリー」すなわち“歴史”とか“経歴”を分析する統計分析手法である。たとえば、下のような表1から、結婚という出来事（イベント）はどのように起こりいつ起こりやすいのだろうか、さらに人によって異なる結婚のタイミングに影響する原因は何かを知りたい時、これを明らかにするのがイベントヒストリー分析である。

では、こうしたイベントヒストリー分析はどのような分野でどのようなデータを分析する際

に用いられているのだろうか。実はその適用例は多様である。表2に示すように、たとえば経済学では、「就職する」ことをイベントとしてとらえ、失業している個々人がどのようなタイミングで職に就いていくか、そのダイナミックなプロセスを分析する研究などがある。社会学

表1 結婚に関するデータ

年 齢	婚姻状態	結婚年齢
29	既婚	25
25	未婚	—
40	既婚	22
32	既婚	25
22	既婚	21
27	未婚	—
39	既婚	26
31	未婚	—
36	既婚	27

注) —の所は、未婚者であるため結婚年齢のデータは得られていない。

\*立命館大学産業社会学部助教授

表2 社会科学分野におけるイベントヒストリー分析の例

研究領域	対象の例	左の例での観測対象
経済学	就職	個人（失業者）
政治学	政権交代	政権、政体
組織生態学	企業の倒産	企業、会社組織
社会移動分析	職業移動までの期間	個人（就業者）
家族社会学	親世帯との同居期間	個人
人口学	死亡、結婚・出産年齢	個人
犯罪学	再犯までの期間	個人（非行経験者）

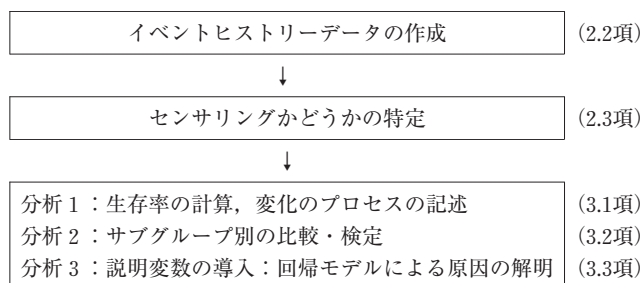


図1 イベントヒストリー分析の分析手順

では職業移動に焦点をあてた例をはじめ、ライフコース上の色々な出来事（たとえば結婚、出産、転職など）を考察する際などにも利用されている。

もっとも、イベントヒストリー分析という呼び方を専ら用いているのは社会学をはじめとする社会科学の分野においてであり、同じ手法が医学・生物学・工学などでは「生存時間分析モデル」「死亡時間分析」「信頼性理論」などとして人・生物・製品・部品・システムの寿命や故障を問題とする際に従来からしばしば用いられている。最も一般的には「生存時間分析」と呼ばれている。

## 2. イベントヒストリー分析の基本的な考え方

### 2.1 イベントヒストリー分析の手順

表2からわかるとおり、私達の周囲には数々

のイベント（出来事）が存在する。これらのそれぞれについては、必ずしも全ての対象が観測（調査）時点で既にイベントを経験しているとは限らない。さらに結婚や就職などのイベントが起こる時点は人によって異なるし、またその起こり方もイベントによって多様である。イベントヒストリー分析の技法がこうした問題にどう対処しているのだろうか。実際の分析手順にそって見ていこう。イベントヒストリー分析の手順は図1のようにまとめられる。なお、各項目の右に（ ）で示す項で、以下に説明を加える。

### 2.2 イベントヒストリーデータの作成

イベントヒストリー分析で扱うデータは、ある個人や個体が、注目するイベントをいつ経験したか、という情報から作られる。

数々の会社について、それらが設立された日

と、倒産もしくは吸収合併 (M & A) され“死亡”した日の情報から、会社・企業の倒産までの存続期間について分析を行うことを考えてみよう。あるいは、初めて職に就いた日から、その初職から他の職に移った (もしくは辞めた) 日までの人々の記録から、初職の勤続期間について分析することを考えてみるのもよいだろう。これらの例では、「倒産」や「(初職からの) 移動」といったイベントが生じるまでの期間 (日数・時間) が分析対象とするデータになる。

### 2.3 不完全なデータへの対処：センサリングの特定

ところで先の例で、前者の場合は倒産・吸収合併されずに生き残っている会社にかんしては、また後者の場合、現在もずっと初職と同じ仕事を続けている人にかんしては、会社の死亡や転職といった注目するイベントが起こるまでの時間がわからないことになる。また、途中で追跡ができなくなって消息不明となり、イベントが起こったのか起こっていないのかわからなくなってしまう例もあるだろう。これらのケースを、まだイベントが起こってないからといって、つまりデータが得られていないからといって無視して分析対象から除外してしまうことはできず、分析の中を含めなければならない。なぜならば、これらのケースを除外すると対象者が偏り、それによって現実の起こり方を反映しないことになるからである。これらのケースについては、「まだ注目するイベントが起こっていない」ことが分析に含めるべき情報なのである。イベントヒストリー分析では、イベントが起こったケースだけではなくイベントがまだ起こっていないケースも同時に分析対象として扱うことができるという利点がある。

とはいうものの、いつイベントが起こるかかわからないのに、したがってイベント生起までの期間がデータとして得られないのに、どのようにして分析に含めることができるのだろうか。イベントヒストリー分析では、これらのケースは「中途打ち切り (センサリング)」と呼ばれる。つまり、社会調査データなどでは、調査を実施した時点までにイベントが起こっていないケースについては、調査以降にイベントが生起する (生起するかもしれない) 個体に関する事象生起時点の情報が得られないまま観測は中途打ち切りとなる。この中途打ち切りという考え方を導入して、イベントが既に起こっているケースか、それとも打ち切り (センサリング) ケースかどうかを識別できるよう、対象ごとに特定することによって、まだ注目するイベントが起こってない打ち切りケースも分析に含めることができるのである。

これらの打ち切りケースについては、打ち切りられた時点までの経過時間が分析対象データとなる。つまり、センサリングと特定されたケースにかんしては、たとえば、まだ生き残っている会社の設立から調査時点までの期間や、あるいは初職を続けている人の就職から調査時点までの期間をデータとして用いる。

### 2.4 生存率・ハザード関数

この項では、イベントヒストリー分析に用いられる統計的概念について若干ふれる。なお、より詳しい概念については、Blossfeld ら (1989) および Tuma ら (1984) を参照されたい。イベントヒストリー分析では、生存率やハザード率という概念が用いられる。生存率とは、ある時点までに、全体の何パーセントがまだイベントが起こらないで残っているかを示す

ものである（通常、イベントヒストリー分析など、一連の生存時間分析では、イベントが起こる前の状態を「生存」と呼ぶので、「生存率」と呼ばれる）。生存率は、観測された生存期間、すなわちイベントが起こるまでの期間についての情報から、打ち切りケースを考慮に入れて計算される。ハザード率とは、ある時点までにはまだイベントが起こらないで残っている者が、そのイベントを経験する単位時間あたりの（瞬間の）率・可能性を示すもので、いわば条件付き確率のようなものといえる。ちなみに、ハザードとは“危険”という意味である。もともとこの統計手法が「死亡」や「故障」の分析に多く適用されたものであったため、注目する事象（イベント）の“おこりやすさ”を、危険を意味するハザードという語を用いて表しているのである。結婚というイベントを例にとれば、24歳の未婚の人が25歳までに結婚する率が、単位時間を1年にとった時のハザード率といえる。ハザード率の考え方をを用いることによって、多様な起こり方をするイベントに対処し、その起こり方を記述したり、またそのイベントの起こり方はどのような要因の影響を受けているのか、複数の説明変数・共変量との関係を明らかにするなどといったことが可能となる。

### 3. イベントヒストリー分析によるデータ解析

イベントヒストリー分析の主な目的は、まず第一にイベントの起こり方の傾向を明らかにすること、第二に、いくつかのグループ間でイベントの起こり方に違いがあるか、生存関数の比較や検定を行うこと、第三に、説明変数を導入した因果モデルを構築し、効果を明らかにすること、などがある。上のような順で分析をすす

めていくのが標準的な方法といえる。

#### 3.1 生存率の計算

例えば「結婚」が時間とともにどのように起こっているのか明らかにする、ということが関心対象であるならば、結婚のしかたはどのようなカーブを描き、いつ頃（何歳頃）最も結婚する可能性が高くなるのか、などを捉えることといえる。このとき、結婚年齢、すなわち未婚であった人が結婚するまでの期間についての情報から生存率の計算をおこなう。さらにこれを、年齢などを横軸に、生存率を縦軸にとって生存関数をプロットしてグラフィカルに示すと、イベントの起こりかたの把握・確認が視覚的に容易になるだろう。生存率（生存関数）の計算方法には、カプランマイヤー法（積一極限法）または生命表法がある。それぞれについてさらに詳しく知りたい場合はBlossfeldら（1989）を参照されたい。

#### 3.2 生存関数の比較・検定

しばしば、対象とするケース全体のイベントの起こり方を把握することにとどまらず、いくつかのグループ間でイベントの起こり方に違いがあるかどうかを比較検討したい、といったことに関心のあることもあるだろう。たとえば、企業規模（大企業か、中小企業か、など）が企業の生存率に影響を及ぼしているのだろうか、あるいは、学歴（大卒か、高卒か、など）によって結婚の起こり方が異なるのだろうか、を調べたいといった場合などである。このとき、大企業、中小企業と零細企業といったグループに分け、グループごとの生存率のプロットを比較したり、さらに生存関数が等しいという仮説を検定することによってこうした問題について検

討を行うことができる。

調べることができる。

### 3.3 説明変数の導入：回帰モデルによる要因の解明

さらに、イベントの起こり方に影響する要因を明らかにしたい場合は、説明変数（共変量）を導入した回帰モデルによる検討を行うことができる。説明変数がイベント時間に与える影響を分析する方法には、1) パラメトリックな回帰モデルによる方法、2) Cox のセミパラメトリックな比例ハザードモデルがある。

#### (1) パラメトリックな回帰モデル

パラメトリックな回帰モデルは通常の（重）回帰分析と似た手法といえる。イベントの起こり方にはどのような要因が関わっているのか、イベントが起こるまでの時間を従属変数とし、影響を与えるであろう変数を説明変数として、回帰モデルを構築する。このときイベントまでの時間がどのように分布しているかの仮定をおこなう。つまり、ハザードの分布形をあらかじめ選択することが必要である。そして、説明変数がイベントの起こり方に与える影響を表す回帰係数の推定を行う。

#### (2) Cox のセミパラメトリックな比例ハザードモデル

生存時間に影響する説明変数の効果を調べるもう一つの方法は、Cox の比例ハザードモデルに基づく回帰分析をおこなう方法である。これはハザード率の分布形をあらかじめ選択することを必要とせず、したがってハザード率の分布が未知の場合でも分析を行うことができる。説明変数の値ごとのハザード率の比（ハザード比）から、一方が他方の何倍起こりやすいかを

### 4. 分析例

#### 例1) 「新しさの脆弱性」仮説の検証

新しい組織は古い組織よりも高い死（倒産、吸収合併）の危険にさらされている、つまり組織の死亡率は組織の年齢に依存する、という「新しさの脆弱性」（スティンチコム、1965）の仮説を検証した研究（Freeman, Carroll & Hannan, 1983）は、イベントヒストリー分析を用いた一つの好例といえるだろう。スティンチコムの議論は以下である、すなわち、新しい組織は新しさの脆弱性の影響を受ける。すなわち、古い組織より死亡のリスクが大きい。なぜなら、低レベルの正統性、確立された組織に対して効率的に競争できない、などの理由からである。さらに、新しい形態の新しい組織は、古い形態の新しい組織より死亡しやすい。

Freeman らは、イベントヒストリー分析の適用によって組織の生態についての検証を行った。

#### 例2) 女性の結婚年齢の分析

女性の結婚のタイミングに影響すると考えられる説明変数の中で、いったい何が重要なのか、およびそれはどのような効果をもつかを、イベントヒストリー分析によって解明したのが以下の結果である（中井、1993）。回帰モデルの分析の結果、学歴や結婚前の職業といった女性の社会的地位は結婚年齢に影響し、高学歴女性ほど、また専門職といったように職業的地位が高いほど、また母親が専業主婦であるほど結婚は遅くなり、その一方で、両親の学歴や父親の職業などは結婚の時期に影響が少ないことが

表3 女性の初婚年齢の分析結果

説明変数	推定値
本人教育	.071**
初職の種類	.166*
初職の官民	-0.021
父教育	0.015
母教育	-0.025
父職威信	-0.003
出身階層との相対的地位	-0.014
母親の就業の有無 (就業 = 1)	-.194**
切片	1.240**
イベント	125
センサーケース	61

注) \*\* P < .01 \* P < .05, 結婚後に初めて職に就いた者を除く。

わかった。このデータでは、イベント (結婚) が起こった人は125人、センサーケース、つまり未婚のため中途打ち切りとなったケースが61人である。ここでは、結婚というイベントのハザード関数としてログロジスティック分布 (初期に急上昇し、以後次第に減少するような形状) を仮定している。

この分析と同様に、結婚および第一子出産年齢のタイミングに注目した他の研究では、いわゆる「専門職」と称される職に就く者の間にも、専門職と準専門職とは明確なライフイベントのタイミングの差異がみられることがイベントヒストリー分析によって明らかにされている (中井, 1998)。

## 5. イベントヒストリー分析を行うための統計パッケージ

大量のデータでより精密な分析を試みる場合はやはり統計パッケージを利用することになるだろう。SAS, SPSS, BMDPなどが、イベントヒストリー分析のプログラムを収録している統計パッケージのなかでも最もよく利用されてい

るものである。たとえば、SASではLIFETESTプロシジャによって生存率 (カプランマイヤーおよび生命表法) の計算, LIFEREGやPHREGプロシジャによって説明変数も導入した回帰モデルを実行することができる (SASの使い方については浜島 (1990), 中井 (1991)などを参照のこと)。またSPSSでは、生存率を計算するSurvivalやKm, 回帰モデルを検討するCoxregがある。

## 6. 主要な文献の紹介

医学統計書での生存時間分析, 工学での信頼性を扱った本は数多くあるが (浜島, 1990など), 社会科学でのイベントヒストリー分析を扱った日本語の簡単な解説書は必ずしも多くない。分析に特徴的な概念・考え方などについて参考になる文献としては, Allison (1984), Blossfeldら (1989) などがある。さらに, Yamaguchi (1991) は分析を行うことを念頭においた構成となっており, 実際の利用の便がはかられている。また, やや難易度は高いが, 多様な分布形などの統計的理論を勉強する際には, Kalbfleischら (1980) が参考になるだろう。Mayerら (1990) には多彩な実証例があり, 分析の参考となる。

## 7. おわりに

イベントヒストリー分析は, イベントが起こるまでの時間を分析対象とし, 観測を打ち切られたデータをも含めた分析を行うことができる。この意味でも非常に有用な分析手法といえる。また, イベントが生起するまでの時間が問題関心の対象であり, そうした経過時間に意味

のある現象であれば分析対象として有効であり、社会科学において広範な利用可能性を持つ。今後、種々の領域での分析応用例が蓄積されることによって、当該領域の研究の発展も期待される。

参考文献

Allison, P. D., 1984, *Event History Analysis: Regression for Longitudinal Event Data*, Sage.

Blossfeld, Hans-Peter, A. Hamerle and K. U. Mayer, 1989, *Event History Analysis: Statistical Theory and Application in the Social Sciences*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cox, D. R. and D. Oakes, 1984, *Analysis of Survival Data*, Chapman and Hall.

Freeman, John, G.R. Carroll and M. T. Hannan, 1983, "The Liability of Newness: Age Dependence in Organizational Death Rate," *American Sociological Review*, 48: 692-710.

Gross, Alan J. and Virginia A. Clark, 1975, *Survival Distributions: Reliability Applications in the Biomedical Sciences*, New York: John Wiley & Sons. 医学統計研究会 (訳) 1984, 『生存時間分布とその応用』海峽書房。

浜島信之, 1990, 『多変量解析による臨床研究』名古屋大学出版会。

Kalbfleisch, J. D. and R. Prentice, 1980, *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.

Mayer, Karl Ulrich and Nancy Brandon Tuma, 1990, *Event History Analysis in Life Course Research*, Madison: The University of Wisconsin Press.

中井美樹, 1991, 「SAS/LIFETEST および SAS/LIFEREG について」『北海道大学大型計算機センターニュース』Vol.23 No.6, 24-41頁。

中井美樹, 1993, 「現代女性の結婚年齢の動学的分析」『現代社会学研究』第6巻, 30-50頁。

中井美樹, 1998, 「女性の職業階層とライフスタイル—専門職女性のライフイベントのタイミング」白倉幸男編『社会階層とライフスタイル』83-100頁。

Tuma, Nancy Brandon and Michael T. Hannan, 1984, *Social Dynamics: Models and Methods*, Orlando, Florida: Academic Press.

Yamaguchi, Kazuo, 1991, *Event History Analysis*, Sage.

数式付録

ここでイベントヒストリー分析に用いられる統計的概念について若干ふれておこう。

(1)生存関数, 確率密度関数, ハザード関数

イベントが起きる時間を  $T$  とする。イベントヒストリー分析では  $T$  の確率分布である生存関数が推定され, さらに確率密度関数, ハザード関数が計算される。

ある時点  $t$  までにイベントが起こらない確率, すなわち生存関数は以下である。

$$S(t) = \Pr[t < T]$$

これは, 時間の経過に伴い  $0$  に近づく, 減少関数である。

逆に, ある時点  $t$  までにイベントが起こっている確率は以下ようになる。

$$F(t) = 1 - S(t) = \Pr[0 < T \leq t]$$

これが死亡関数である。

したがって,  $T$  の確率密度関数  $f(t)$  は,  $F(t)$  の微分によって表すことができ,

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t < T \leq t + \Delta t]}{\Delta t} = \frac{-dF(t)}{dt}$$

となる。さらにハザード関数は, 「 $t$  時点まではイベントが生起していない者のなかから, 新たに,  $T = t$  の瞬間にイベントが起こる率」を表し, 以下のように定義される。

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t < T \leq t + \Delta t | T > t]}{\Delta t}$$

(2)パラメトリックな回帰モデル

センサリングを含むデータにパラメトリックなモデルを仮定し, 共変量を投入したモデルの活用により種々の要因の影響を明らかにする方法で, SAS では LIFEREG プロシジャにより行う。

イベントが生起するまでの時間を  $T$  とし, 共変量 (説明変数) を  $x' = (x_1, \dots, x_n)$ ,  $\beta = (\beta_1, \dots, \beta_n)$  が推定すべき回帰パラメータのベクトル, とする。このモデルで, 生存時間に対する共変量の効果は乗

法的と仮定され、したがって基準分布を  $T_0$  とすれば、共変量  $x$  を持つ個体の生存時間  $T$  は以下で表される。

$$T = \exp(\beta x) T_0$$

従属変数となるイベントの生起時間  $T$  は通常対数変換が行われ、 $x$  の影響は加算的に表現される（すなわち、 $T$  に関しては乗法的に影響する）という線型の

モデルで表すことができる。したがって、

$$\log T = \beta x + \log T_0$$

通常この  $\log T_0$  には切片パラメータと尺度パラメータが含まれ、切片を  $\mu$ 、尺度パラメータを  $\sigma$  とよび、イベントについてどのような分布が想定されるかによっていろいろなモデルについて適用できる。



## Statistical Models and Methods for Life Event Data in Social Science Research: Event history analysis

NAKAI Miki \*

**Abstract:** Event history analysis, or survival analysis, was originally developed to analyze data from medical trials and industrial equipment tests data. However, it is now widely used for describing the likelihood of events in various social science areas. The purpose of this article is to provide an introduction to the statistical modeling and applications of event history analysis. It also explains the basic concepts associated with the analysis of censored data. Some examples are also given of how event history (survival) data can be analyzed using these techniques.

**Keywords:** event history analysis, longitudinal event data, censored data, survival function, timing of marriage

---

\* Associate Professor, Faculty of the Social Sciences, Ritsumeikan University