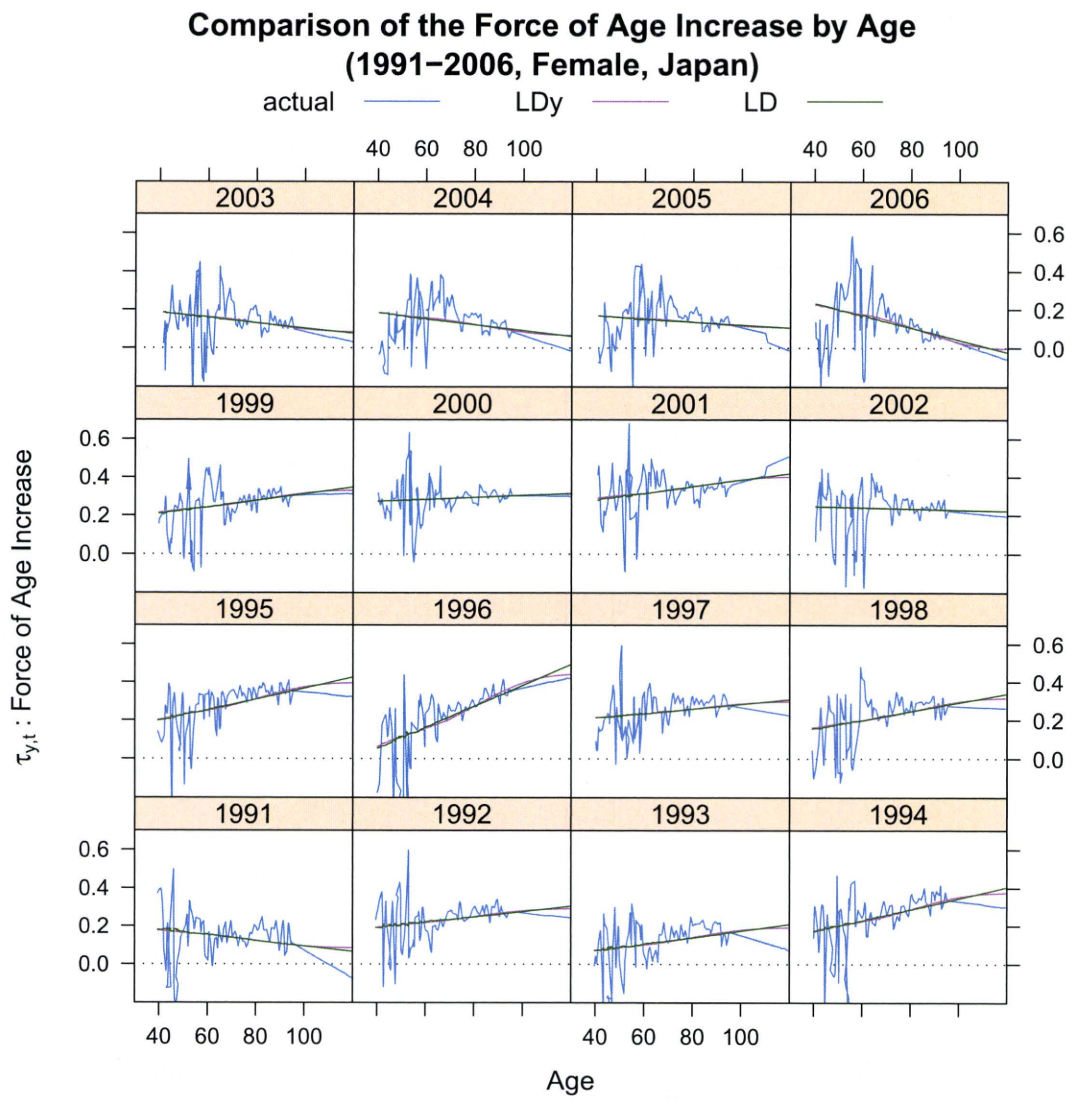


図 6 Comparison of the Force of Age Increase by Age(1991-2006, Female, Japan)



3 統計学的観点からのモデル比較

昨年度の研究において、高齢死亡率改善を死亡率の「低下」として捉えるべきなのか、「シフト」として捉えるべきなのかという問いに定量的な回答を与えることを目的として、「低下」型モデルと「シフト」型モデルの両者のパフォーマンスを統計学的観点から比較を行った。「低下」型モデルで最もあてはまりがよかったのは LC モデル、「シフト」型モデルでは LD モデルであったところから、この 2 モデルを用いて比較・検討を行ったが、本年度の研究において改良された LD モデルにおいても同様の結論が得られるかについて検証を行ってみることとしよう。

用いるアプローチは昨年と同様、以下の通りである。

1. 真の死亡率がモデルによって推定された率であると仮定する。
2. 死亡数は二項分布 $B(N_{x,t_c}, p_{x,t_c})$ に従うものとする。ただし、 N_{x,t_c} 、 p_{x,t_c} は、 x 歳、暦年 t_c における人口及び死亡率を表す。
3. N_{x,t_c} は、死亡リスクに曝される生存述べ年数 $E_{x,t}$ に最も近い整数により近似する。

昨年度同様、日本女性人口の N_{x,t_c} が非常に大きい値であり、通常信頼区間の算定で用いられるような危険率では多くの実績値が信頼区間の外に出て、有効な比較とならない可能性があることから、信頼区間の算定にあたっては、0.01% という極めて低い危険率を設定することとした。

図 7、8 は対数死亡率の実績値 (“actual”で示されたもの) 及びその (100 - 0.01)% 信頼区間 (“upper”と“lower”で示されたもの) と、LC モデルによる対数死亡率との差を見たものである。一方、図 9、10 は同じ図を LD モデルについて観察したものである。これらの傾向は、昨年度の改良前の LD モデル (LDy) を用いた場合と概ね同様であり、両者を比較してみると、LC モデルについては、80 歳近辺において、多くの実績値が信頼区間の外に出てしまっている。一方、LD モデルについては実績値が外に出ている数はより少ない。したがって、改良された LD モデルについても、LC モデルよりもパフォーマンスが優れているのではないかということが示唆される結果となった。

また、これをより明確に見るため、図 11 に、各年齢について対数死亡率の実績値が信頼区間の外側に出る割合を示した。これによれば、高い年齢で LD モデルの方が割合が高い部分は若干あるものの、総じて LD モデルのパフォーマンスが LC モデルをかなり上回っていると見る事ができるだろう。この両モデルの比較の結果は、昨年度同様、わが国の近年の高齢死亡率改善について、「低下」よりは「シフト」と捉える見方がより強く支持されることを示唆しているものと見る事ができる。

図7 Difference of Log Actual Rate and CI against Model (LC)(1970-1989, critical value = 0.01%)

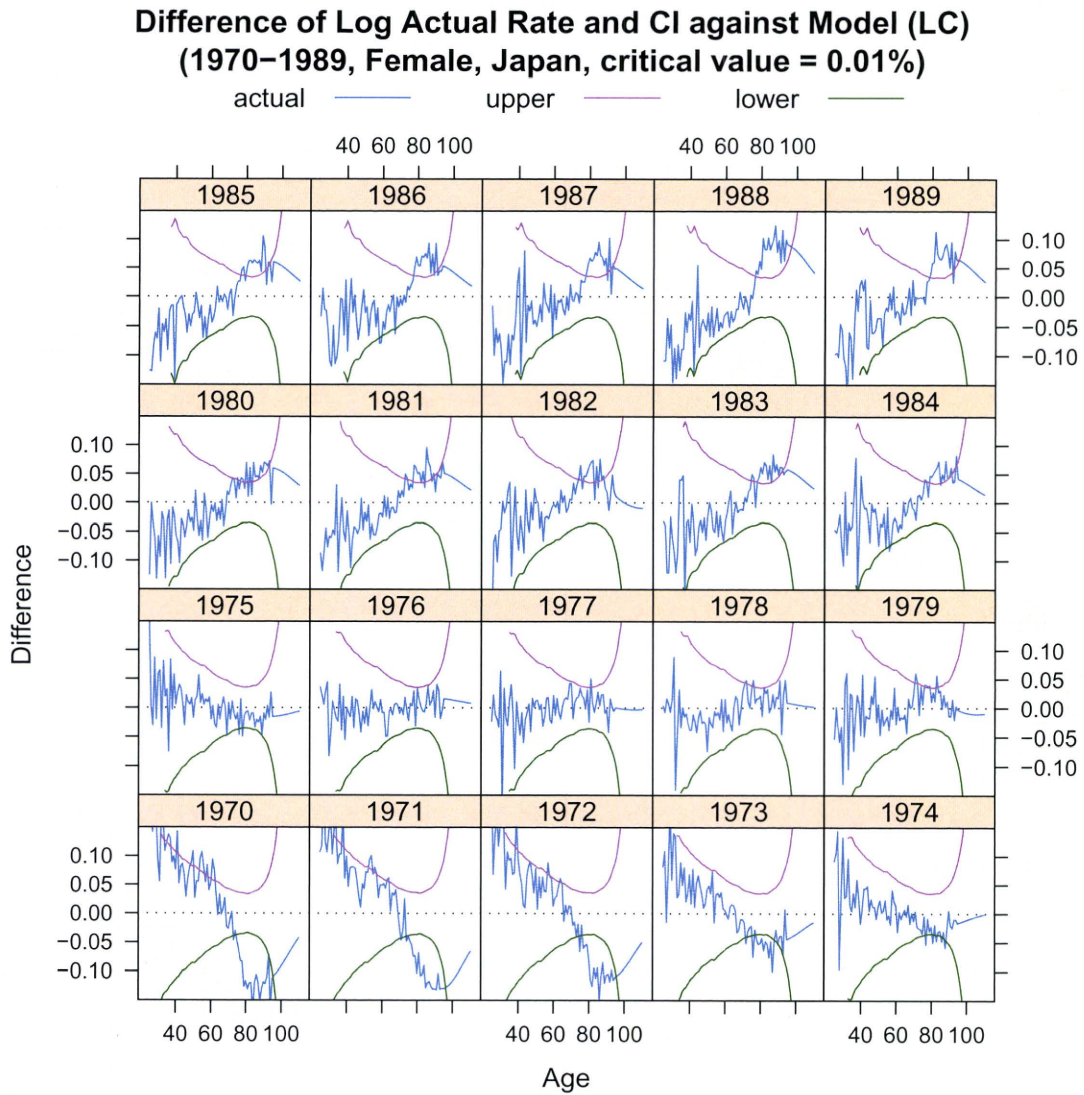


図8 Difference of Log Actual Rate and CI against Model (LC)(1990-2009, critical value = 0.01%)

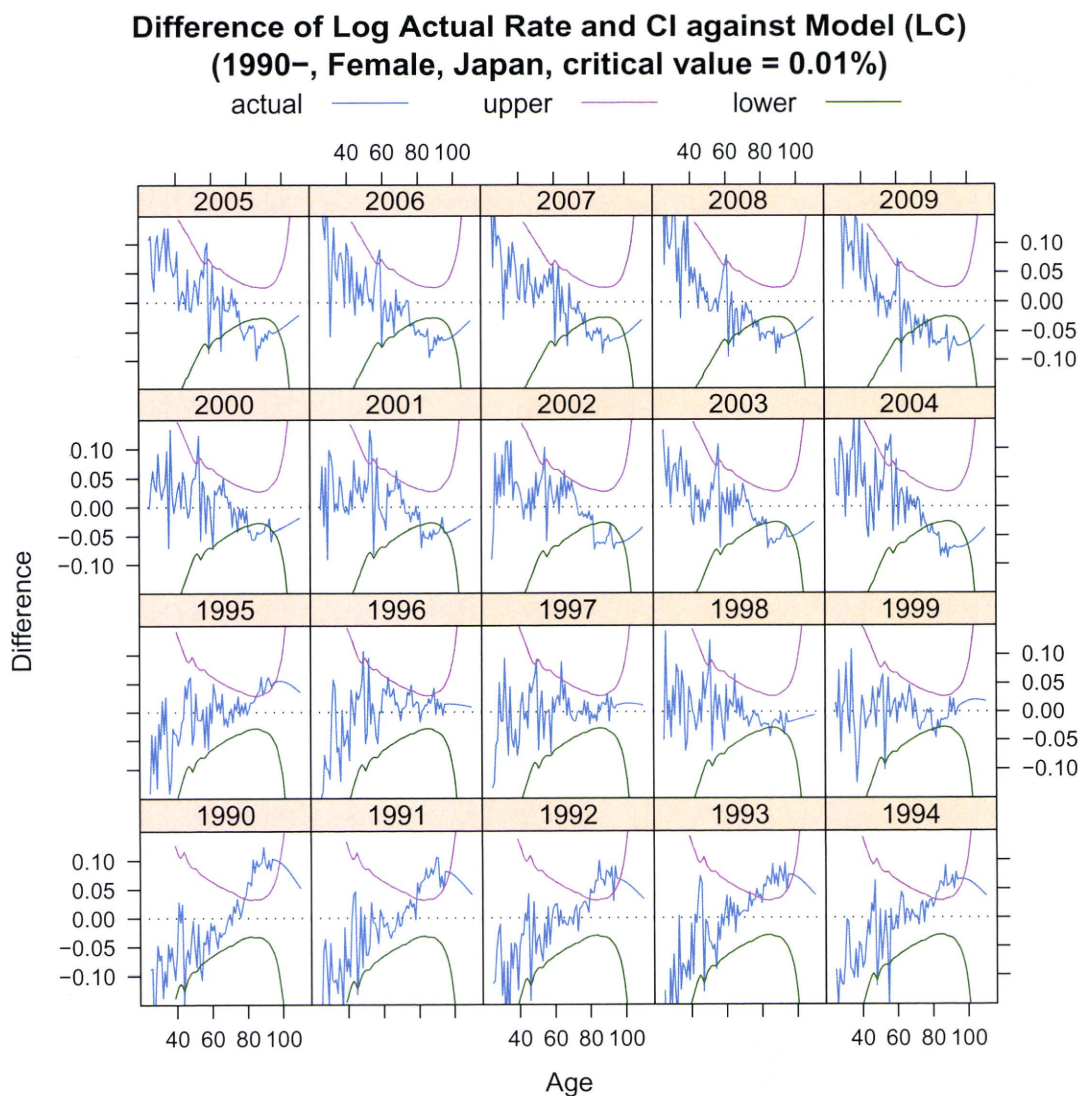


図9 Difference of Log Actual Rate and CI against Model (LD)(1970-1989, critical value = 0.01%)

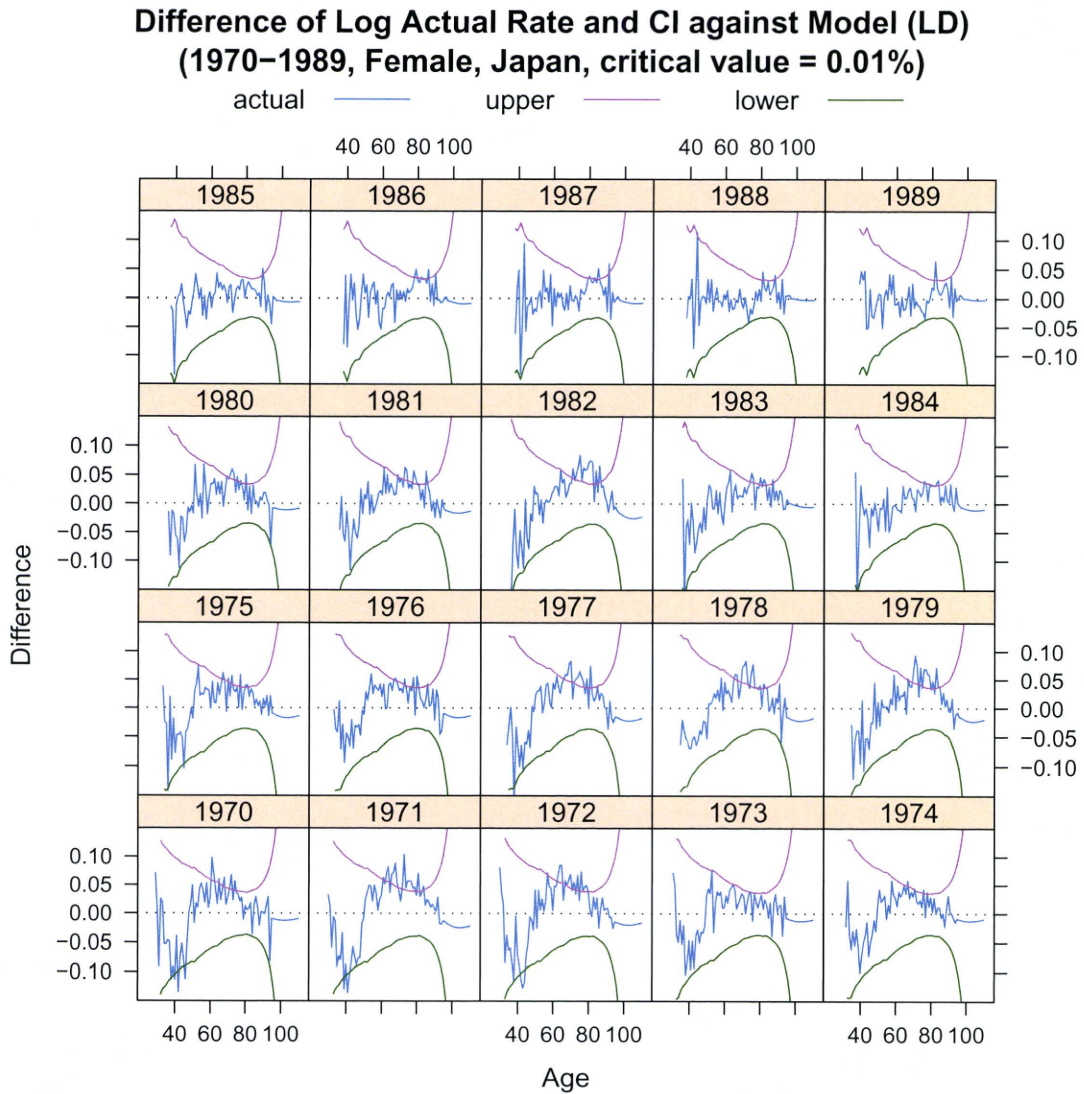


図 10 Difference of Log Actual Rate and CI against Model (LD)(1990-2009, critical value = 0.01%)

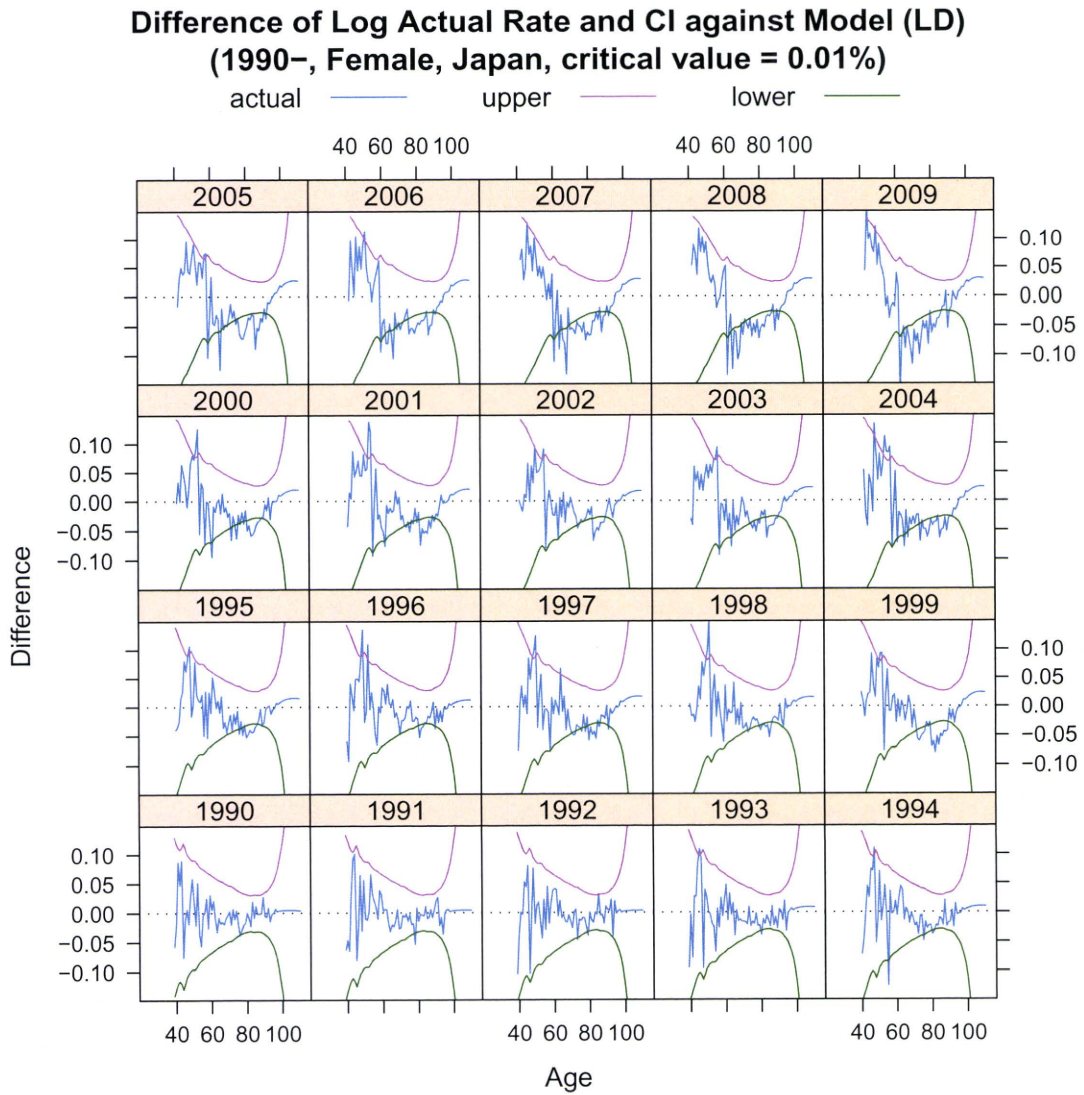
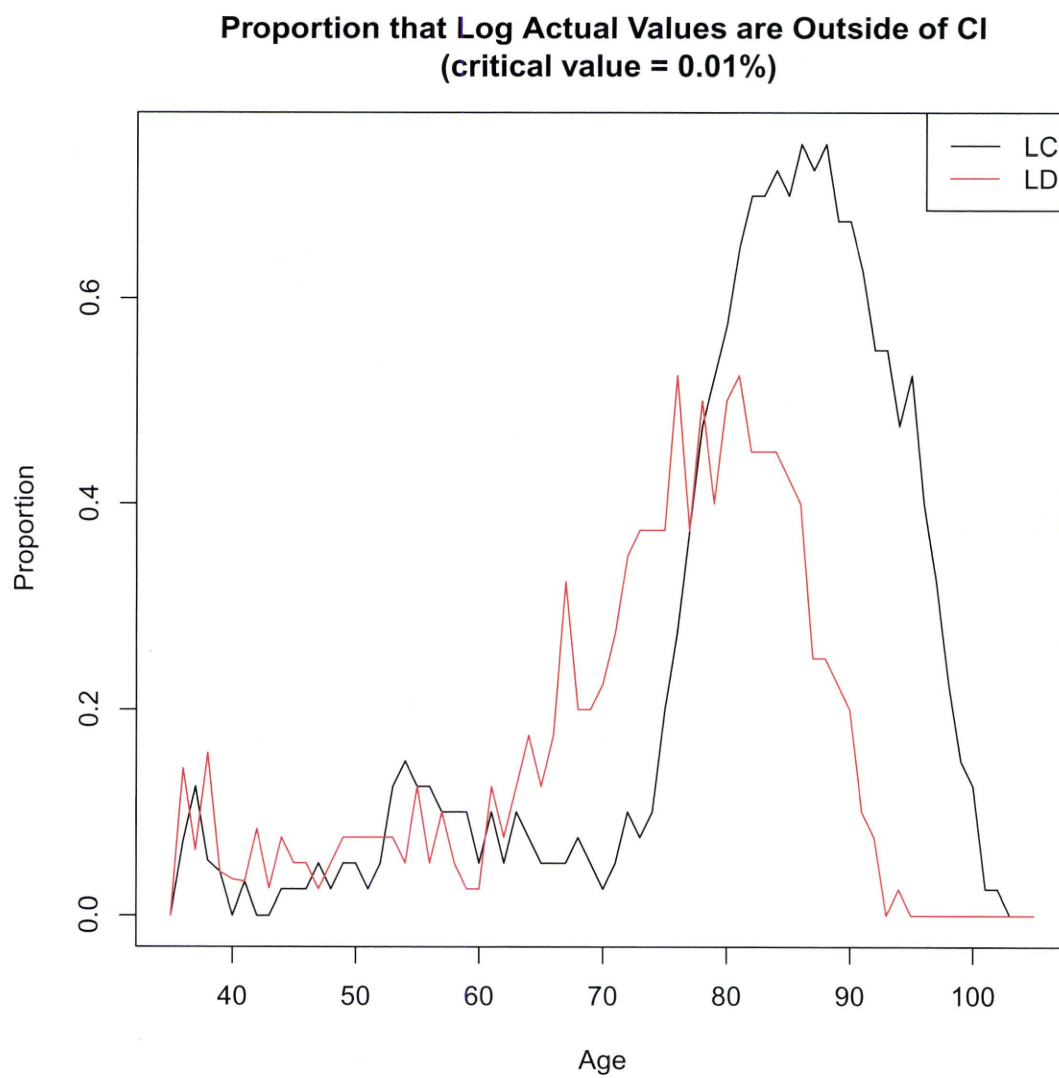


图 11 Proportion that Log Actual Values are Outside of CI (critical value = 0.01%)



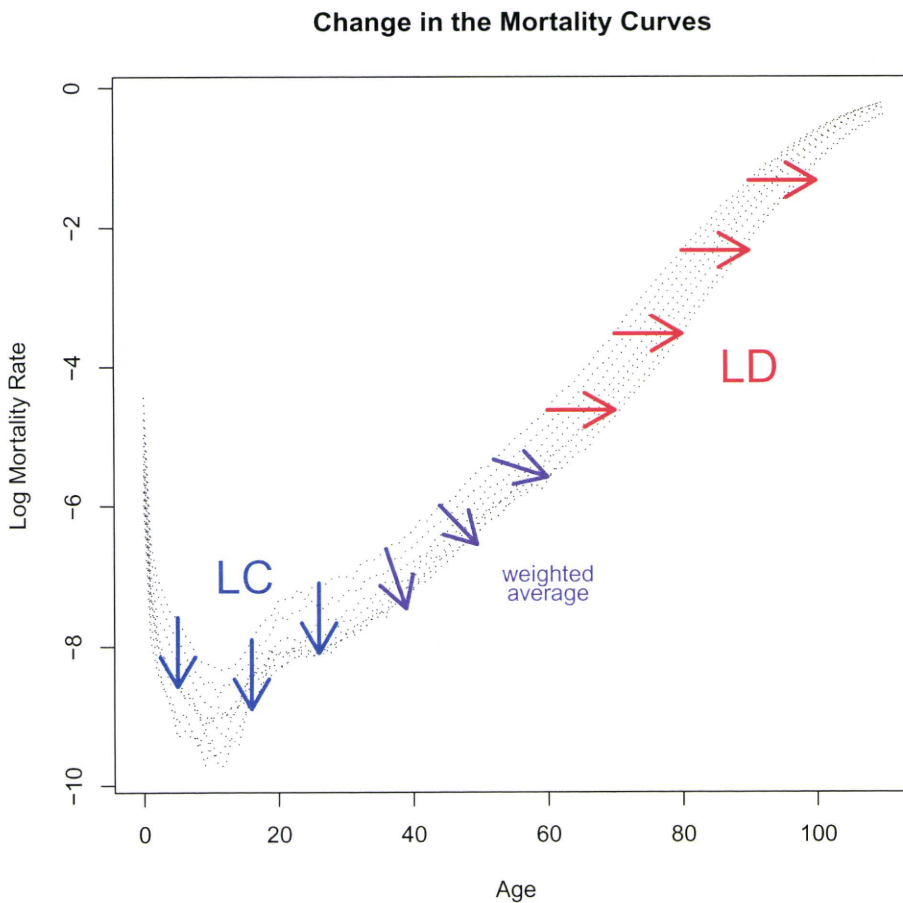
4 対数死亡曲面上の接ベクトル場

4.1 対数死亡曲面上の接ベクトル場を利用した全年齢死亡モデルの構築

前節までにおいて、高齢者死亡のモデリングにおける LD モデルの有効について、LC モデルとの比較を通じて明らかにしてきた。しかしながら、LD モデルは成人死亡率のみを表現するモデルであり、そのまま全年齢死亡率モデルとして利用することはできない。

そこで、本節においては、高齢部分に LD モデル、若年部分に LC モデルの構造を持つ全年齢死亡率モデルを、対数死亡率曲面 S 上の接ベクトル場を利用して構築する方法論について検討する。まず最初に、模式的な例によってこのような全年齢死亡率モデルの構成法を考えることとしよう。図 12 は対数死亡率の改善例を表したものである。我々はこれまでの検討によって、高齢部分の死亡率改善は、赤い矢印で示された方向に死亡率改善を表現する LD モデルがよくあてはまる一方、若年部分については青い矢印で示された方向への改善で表現する LC モデルがよくあてはまることを結果として得た。ここで、これらの

図 12 Change in the Mortality Curves



矢印は、対数死亡率曲線上のそれぞれの点において、その時刻にどちらの方向に死亡率が改善しようとしているかを表している。数学的には、これらの矢印が表現するものを、さらに時間軸を加えた対数死亡率曲面上における接ベクトル場として定式化することが可能である。

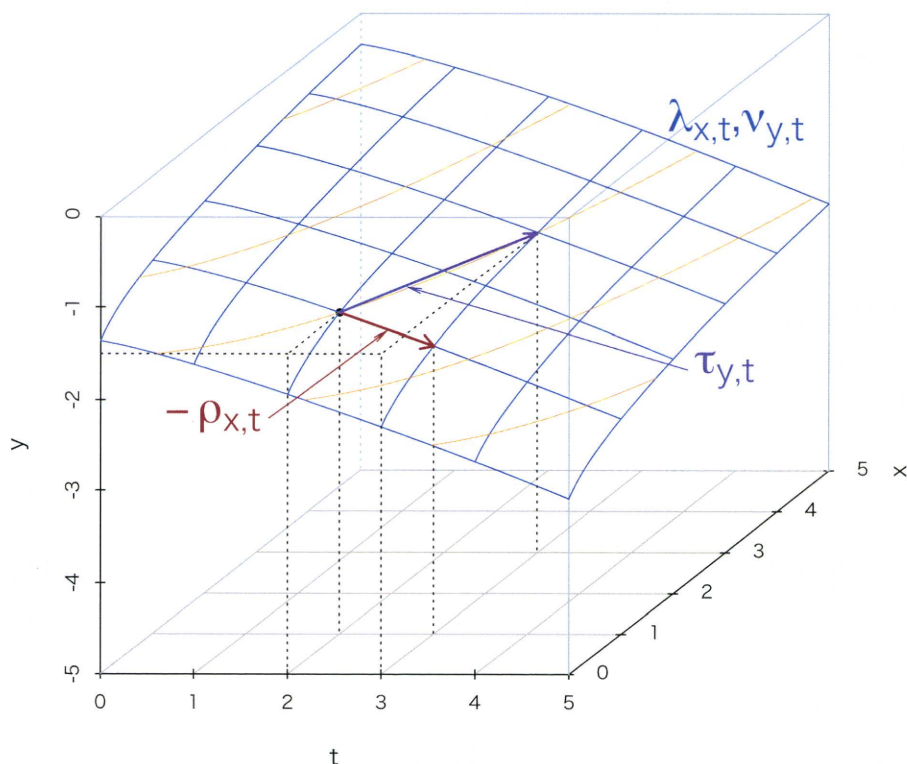
昨年度の研究において、死亡モデルを対象となる関数の時間変化である微分型表示で表し、そこで用いられる関数 $\rho_{x,t}$, $\nu_{y,t}$ を用いたベクトル

$$\rho(x_0, t_0, y_0) = (0, 1, \rho_{x_0, t_0})$$

$$\tau(x_0, t_0, y_0) = (\nu_{y_0, t_0}, 1, 0)$$

が、 S の接ベクトルになることを論じた。そして、 S の各点にこの接ベクトルを対応させることにより、 S 上に 2 種類の滑らかな接ベクトル場を定義することができた。すなわち、LC モデルに対応する接ベクトル場と LD モデルに対応する接ベクトル場である。こ

図 13 Log Mortality Surface and Two Differential Functions



れら各点における接ベクトルは、LC, LD モデルのパラメータ推定を行うことにより求め

ることが可能である。LC, LD モデルに対応する 2 つの接ベクトル場を推定した上で、図 12 における紫の矢印で表される中間的な年齢層に対しては、両者のベクトル場に属するベクトルを、一定のウエイトを用いてベクトルレベルで加重平均することにより、LC モデルから LD モデルへと連続的につながるようなベクトル場を定義することが可能である。ここでは、ウエイト $w(x, t)$ を線形関数を用いて以下の式で定義することとした。

$$w(x, t) = \begin{cases} 0 & (x < x_0(t) = 25) \\ \frac{x - x_0(t)}{x_1(t) - x_0(t)} & (x_0(t) \leq x < x_1(t)) \\ 0 & (x \geq x_1(t)) \end{cases}$$

ただし、 $x_1(t)$ は、基準年 ($t_{bp} = 1989.5$ 年) において $x_1(t_{bp}) = 50$ とし、その他の年については、 $x_{t=50, t_{bp}}$ を満たす x を $x_1(t)$ とした。そして、これを用いて、

$$(1) \quad w(x, t)\rho(x, t, y) + w(x, t)\tau(x, t, y)$$

が各点における死亡率改善を示すベクトルとなる接ベクトル場を持つような対数死亡率曲面を構成することにより、全年齢死亡モデルを定義することができることになる。この様子を三次元プロットで表したものが図 14、さらにこれを二次元平面に射影したものが図 15 である。

この接ベクトル場による死亡モデルを用いた死亡率推計結果を 1970, 1980, 1990, 2000, 2009 年について示したものが図 16 である。青い点線で示したものが LC モデルによる推計値、赤い点線で示したものが LD モデルによる推計値、黒い実線が接ベクトル場による死亡モデル (以下、「TVF モデル」と呼ぶ) による推計値である。これを見ると、TVF モデルによる推計結果は、25 歳未満の若年層では LC モデルに一致し、それ以降 LD モデルのラインに近づき、高齢層では LD モデルに一致していることがわかる。したがって、対数死亡率曲面上の接ベクトル場を利用することにより、高齢部分では LD モデルによるシフト型の死亡率改善をしつつ、若年部分においては LC モデルによる低下型の死亡率改善をするような、自然なモデリングを行うことが可能となった。

图 14 示例构造一个切向量场

Example of the Tangent Vector Fields

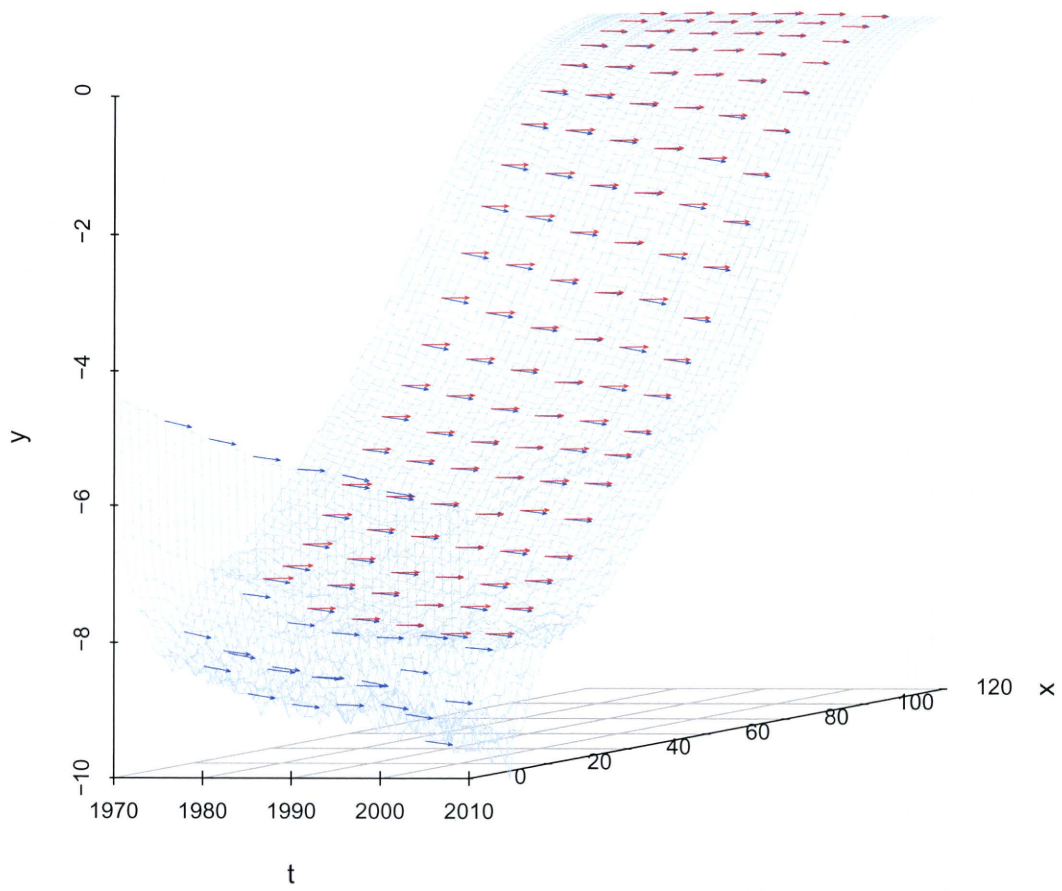


图 15 Example of a Construction of a Tangent Vector Field

Example of a Construction of a Tangent Vector Field

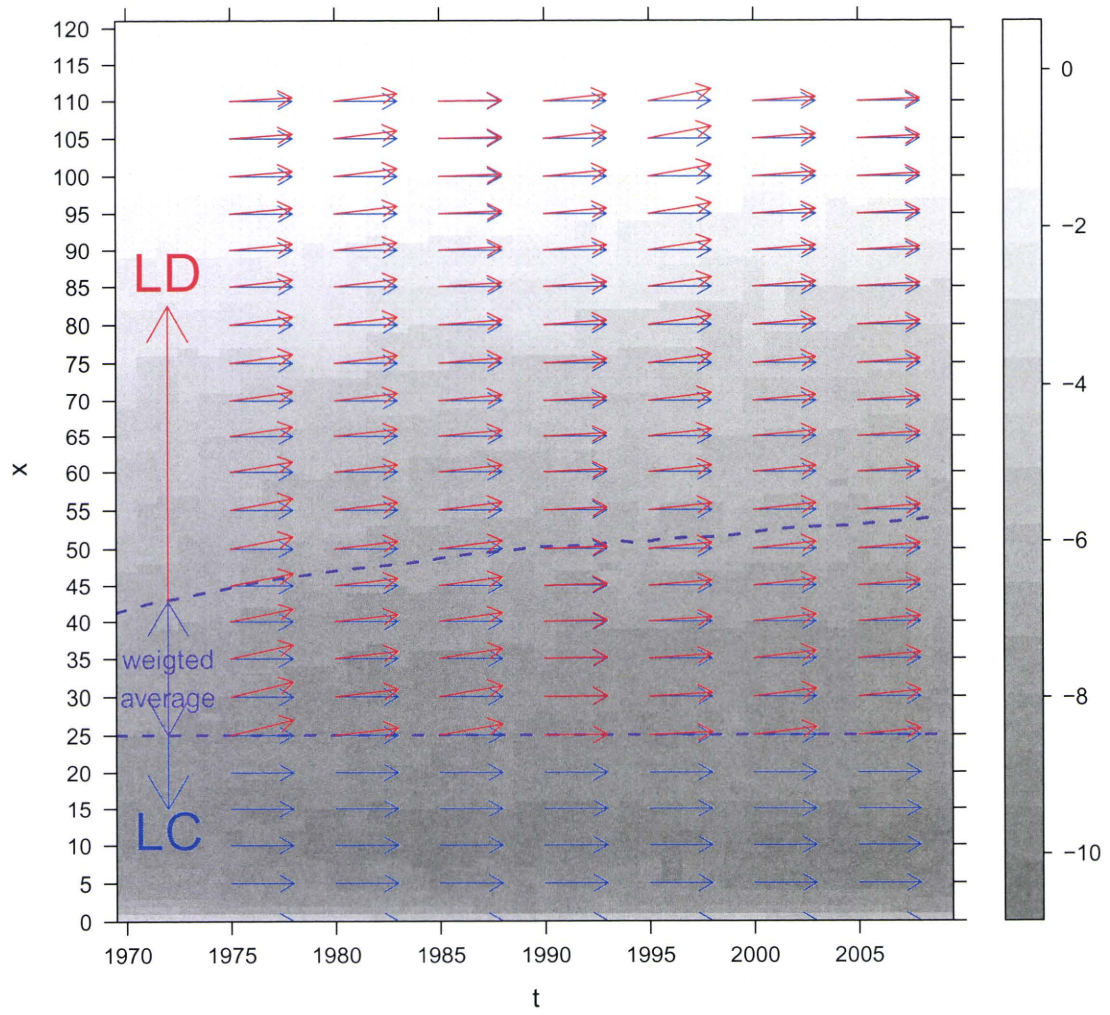
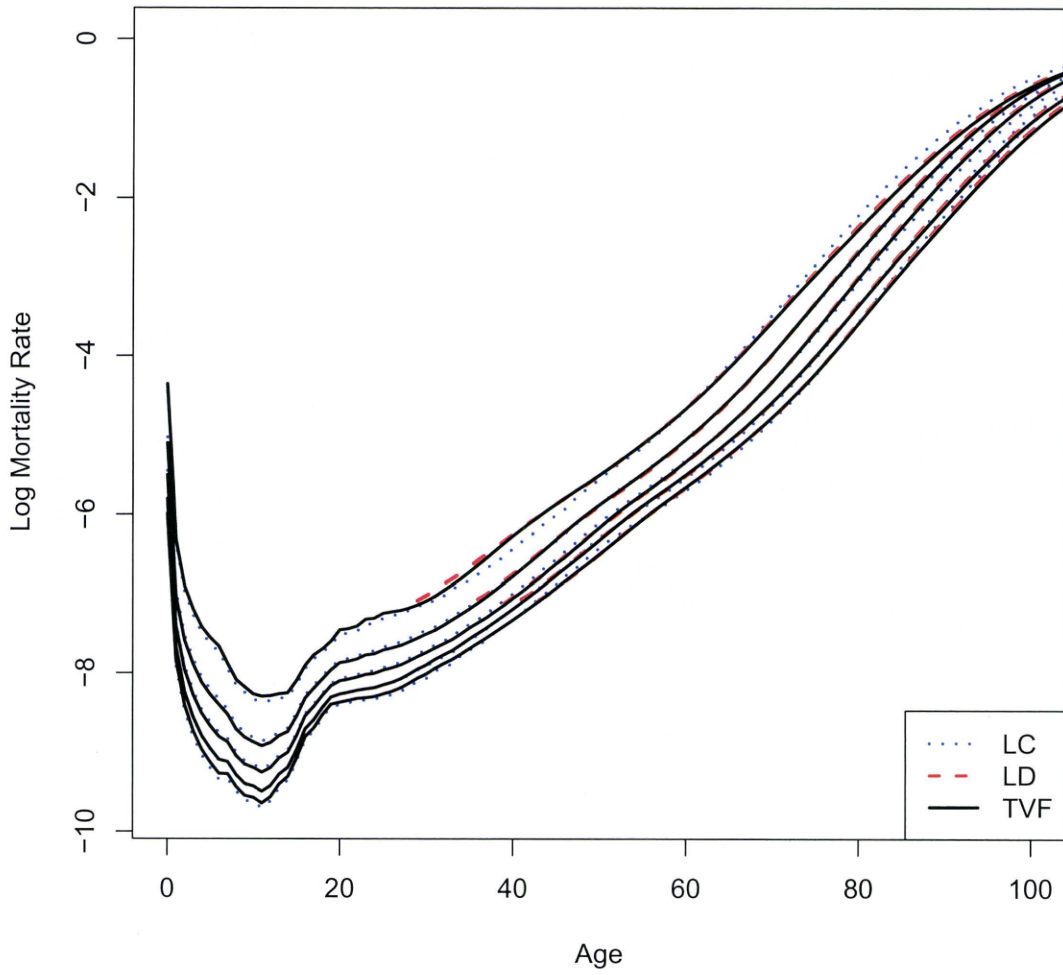


図 16 Estimated Mortality Rates (LC, LD, TVF)

**Estimated Mortality Rates by LC, LD, TVF models
(Female, 1970, 1980, 1990, 2000, 2009)**



おわりに

本年度の研究では、昨年度までの研究成果を発展させ、LD モデルについて、ロジスティックモデルとより整合性を高める観点からの改良を行うとともに、死亡率のモデル化にあたって、対数死亡率曲面上の接ベクトル場を利用することの有効性を示し、高齢部分に LD モデルの構造を持つ全年齢死亡率モデル構築の方法論に関して検討を行った。

昨年度の研究において、シフト型で死亡率改善を表現する「LD モデル」を提案し、現在、標準的な死亡率推計モデルとして用いられているリー・カーター・モデルとの比較を通じてこのモデルの有効性を示したところであるが、LD モデルは年齢シフトモデルのシフト量・勾配パラメータ等を求める際の理論的基礎となっている 2 パラメータロジスティックモデルとの整合性の観点で課題があった。本年度の研究ではこの点を改良し、2 パラメータロジスティックモデルの $y_{i,t}$: 年齢増加率が持つ特性を保有するよう、 $y_{i,t}$ が x の線形関数であるとして LD モデルを定義することとした。この改良された LD モデルに対しても、昨年度行ったのと同様に統計学的観点からの LC モデルとの比較を行い、やはりわが国の近年の高齢死亡率改善について、「低下」よりは「シフト」と捉える見方がより強く支持されることを示唆しているものと見る事ができるとの結論を得た。

また、この LD モデルは成人死亡率のみを表現するモデルであり、そのまま全年齢の死亡率モデルとして利用することはできない。そこで、対数死亡率曲面 S 上の接ベクトル場を利用することにより、高齢部分に LD モデル、若年部分に LC モデルの構造を持つ全年齢死亡モデルである TVF モデルの開発に関する検討を行った。これは、死亡モデルを対象となる関数の時間的変化である微分型表示を利用し、LD, LC モデルを対数死亡率曲面上の接ベクトル場として捉え、中間的な年齢層では接ベクトルのレベルで LD, LC モデルに対応する接ベクトルを加重平均することにより両モデルを接合して全年齢死亡モデルを構成するものである。TVF モデルによる推計結果は、25 歳未満の若年層では LC モデルに一致し、それ以降 LD モデルのラインに近づき、高齢層では LD モデルに一致している。このように、対数死亡率曲面上の接ベクトル場を利用することにより、高齢部分では LD モデルによるシフト型の死亡率改善をしつつ、若年部分においては LC モデルによる低下型の死亡率改善をするような、自然なモデリングを行うことが可能となった。

本研究で検討を行った TVF モデルを、年齢シフトモデルと比較すると以下のような利点がある。年齢シフトモデルで用いられている年齢変換を行うためには、まず最初に死力曲線にロジスティックモデルをあてはめ、パラメータ推定をする必要がある。これは、年齢変換後に行うリー・カーター・モデルによるモデリングがノンパラメトリックな手続きによっているのに対し、年齢変換の中ではパラメトリックな手続きが必要であることを意味しており、一つのモデルの中に両者の手続きが混在していたことになる。しかしながら、TVF モデルでは、このような手続きは必要なく、モデル全体を完全にノンパラメト

リックな手続きのみで構築することが可能となる。したがって、TVF モデルは、高齢者死亡の表現性にすぐれ、年齢シフトモデルの理論的根拠ともなっていたロジスティックモデルの特性を保ちつつ、モデル内部の整合性に関してもより高いレベルで達成することができるのである。

以上に述べてきた通り、TVF モデルは年齢シフトモデルで開発されてきた、シフト型の死亡率改善をモデルに取り込むという観点を採り入れつつ、全体をノンパラメトリックな手続きで構成するという LC モデルが持つ特性を同時に達成したモデルであるということが出来る。この新たなモデルをわが国の死亡動向の分析に活用するとともに、死亡率の将来推計に応用していくことが今後の課題といえよう。

参考文献

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

Ishii, F. (2008) “Mortality Projection Model for Japan with Age-Shifting Structure”, Paper presented at 2008 Annual Meeting of Population Association of America (New Orleans).

石井太 (2008) 「近年のわが国の死亡動向に適合した 将来生命表推計モデルの研究-年齢シフトモデルの開発-」, 『人口問題研究』, 第 64 巻, 第 3 号, pp.28-44.

Ishii, F. (2010) “Application of Age-Transformation Approaches to Mortality Projection for Japan”, Paper presented at Joint Eurostat/UNECE Work Session on Demographic Projections, Lisbon, Portugal.

国立社会保障・人口問題研究所 (2007) 『日本の将来推計人口—平成 18 年 12 月推計—』, (財) 厚生統計協会.

Lee, R. and L. Carter (1992) “Modeling and Forecasting U.S. Mortality”, *Journal of the American Statistical Association*, Vol. 87, No. 419, pp. 659-675.

Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998) *The Force of Mortality at Ages 80 to 120*: Odense University Press.

3 対数ガンマ分布（コール・マクニール分布）の多段階モデルの性質を 利用した結婚・出生に関する行動モデルの開発

Multistage Models of Marriage and Birth: An Extension of the Coale-McNeil Nuptiality Model

金子 隆一
三田 房美

要 約

コール・マクニール結婚モデル（CMモデル）は、初婚や出生順位別出生の年齢別スケジュールを与える以外に、多段階より構成される行動のモデルとなっている。この性質を利用することにより、帰結としての初婚・出生年齢分布からその前事象の年齢分布や前事象経験者数を推定することができ、こうした行動的解釈によりモデルの予測性が向上する可能性があることから、CMモデルの多段階モデルとしての形式を開発し、パラメータ推定の統計的手続きを整備した。具体的には対数ガンマ分布を多段階行動モデルとして再定義し、わが国の実地調査データ、人口動態統計データに当てはめた結果、第2子年齢別出生率から第1子年齢別出生率を一定の確度での推定が可能であることが示された。これを応用するとたとえば、第1子年齢別出生率から年齢別婚姻率や第1子出生のリスク人口の推定などに応用され、同棲・事実婚などによる配偶関係統計のかく乱や婚前妊娠・婚外子など出生との関係のかく乱が有る状況下での行動分析、将来推計に用いることができる。また、こうした行動モデルの開発は、少子化をもたらしている要因の解明にも必須であり、有効である。

Multistage Models of Marriage and Birth: An Extension of the Coale-McNeil Nuptiality Model¹

Ryuichi Kaneko² Fu sami Mita³

Abstract

In this study, I first focus on the potential profile of the Coale-McNeil nuptiality model (CM model) as a multistage model of vital events such as the first marriage and birth separated by birth order. The convolution structure of the CM model can be regarded as an expression of a multistage process consisting of attainment of marriageable age, and several waiting time to the goal. This interpretable nature of the CM model is carefully examined here and the corresponding equivalent convolution model is presented. The critical problem is that the assumption of the model over independence between the sub-processes is not satisfied in the reality according to my survey results. Hence I proposed new model by introducing liner relationships among parameters so that realistic associations of the sub-processes be reproduced. It is applicable to describe latent processes such as age at becoming marriageable and marriage market population, for instance.

Introduction

The marriage model developed by Coale and McNeil (1972) provides an elegant example of a multistage model that might be applied to a wide range of events. Their model is a limiting probability distribution of the convolution of infinite numbers of related exponential distributions, and can be regarded as the convolution of a uni-modal distribution of its own form and some number of exponential distributions. Therefore it provides a general mathematical model applicable to multistage processes, by which I mean a process that consists of multiple processes whose completion is required for the target event to occur⁴. Unfortunately, this aspect of the CM model has drawn only limited attention in contrast to its popularity as a standard schedule of first marriage. In this paper, I focus on this profile of the CM model, and examine its potential ability as a behavioral multistage model of the first marriage process.

¹ This is an early version of the study report, or an extended abstract of the paper to be presented to The Annual Meeting of the Population Association of America (Dallas, TX, April 15-17, 2010).

² National Institute of Population and Social Security Research, Japan. e-mail: r-kaneko@ipss.go.jp

³ National Institute of Population and Social Security Research, Japan. e-mail: f-mita@ipss.go.jp

⁴ The multistage model in this context is a special case of the multistate model in which return transitions to previous states are prohibited.

Multistage Framework of First Marriage Process

Coale and McNeil viewed first marriage as a multistage process in which entry into marriageable state, meeting of the eventual spouse, and engagement are required to take place prior to the marriage⁵ (Coale and McNeil 1972). They proposed convolution models to formalize this view of the process. Their model for the first marriage schedule is viewed here as a behavioral model of the process as well, since it is composed of a convolution of multiple distributions which can individually be interpreted behaviorally. Coale and McNeil also proposed a model consisting of a convolution of a normal and some exponential distributions solely for extracting a behavioral explanation. Before examining these models, I first formulate a framework in which to study the multistage aspect of the first marriage process.

Let the following random variables be components of the first marriage process;

- Z : age at first marriage
- X_0 : age at entry into marriage market (age at becoming marriageable)
- X_1 : age at first encounter with eventual spouse
- X_2 : age at engagement
- T_1 : waiting time from entry into marriage market to encounter
- T_2 : waiting time from encounter to engagement
- T_3 : engagement period

Then the following relationships hold among these variables in a first marriage process:

$$\begin{aligned} Z &= X_0 + T_1 + T_2 + T_3 \\ &= X_1 + T_2 + T_3 \\ &= X_2 + T_3 \end{aligned} \tag{1}$$

According to this formulation, the process starts with entry into marriage market and goes through three waiting periods by the time of marriage. Thus the age at first marriage is defined as the end point of the process. The second and third lines of the equation indicate that the beginning of the process may be age at first encounter with spouse, or age at engagement.

Entry into the marriage market is merely a time point at which a person becomes eligible to marry. Physical maturity, arrival at the legal age for marriage, satisfaction of social and economic requirements such as completion

⁵ They acknowledged Griffith Feeney for initial inspiration of this view.

of school or entering employment, and psychological readiness for marriage are all considered as necessary conditions of the entry. Since the specific events that cause a person to enter the marriage market vary widely from person to person and may well be largely psychological, this notion may be best viewed as an abstraction to assist in description of the model.

According to Coale and McNeil (1972), X_1 should be the time of “meeting or stating to keep frequent company with or beginning to date the eventual” spouse. However, I employ “meeting with eventual spouses” as the event for X_1 due to ease of observation⁶.

In order to develop a specific demographic model out of this basic framework, I should identify probability distributions for each variable and the relationships among them. The first attempt to assemble a multistage model should be conducted via the convolution of distributions that are allocated to sub-processes.

The convolution provides the distribution of the sum of two random variables, assuming they are mutually independent. Let X and T denote two independent random variables, which have PDFs, $f(x)$ and $h(t)$ respectively. Then the PDF of $Z=X+T$, $g(z)$ is given by⁷:

$$g(z) = \int_{-\infty}^{\infty} f(x)h(z-x)dx. \quad (2)$$

For instance, suppose that X is age at entry into marriage market and T is waiting time until marriage, at which point age is Z . Then distribution of age at marriage is given by $g(z)$.

Some convolutions of standard distributions are well known. For instance, the convolution of two normal distributions is another normal distribution with sum of the original means as the new mean, and sum of the original variances as the new variance. The convolution of multiple identical exponential distributions is a gamma distribution with two parameters, one of which is the number of exponentials, and the other is the parameter of the exponential distributions. I should note that the independence of the random variables is required when applying convolution.

The framework described here is applicable to many compound events. Birth is one example in which several events are required to take place prior to the event, such as marriage or union formation, onset or resumption of

⁶ A drawback of this specification is that meetings in childhood in such a case as marriage between neighbors are included. Those meetings occur much earlier than entry into the marriageable state. But according to our survey described later, such meetings are exceptional in present-day Japan (1.5% in '97 survey). Further, it is possible to eliminate them based on information regarding the type of meeting.

⁷ If $T > 0$ holds, then the upper boundary of the integral should be z instead of plus infinity. This is the case when T stands for waiting time.

ovulation, copulation, and conception. Some authors have employed multistage framework with convolution models to analyze birth interval (D' Souza 1974, Zhu 1994, and Liang 2000).

The Coale-McNeil Distribution as a Convolution Model

Coale and McNeil (1972) proposed two different types of convolution models for the first marriage process; (1) a distribution that is a limiting distribution from the convolution of infinite numbers of related exponentials with means in a harmonic sequence, and (2) a distribution given by the convolution of the normal distribution and several exponential distributions. First I examine the form and properties of model (1) here. Model (2) is discussed in the following section.

In terms of probability density function (PDF), the model (1) is given by:

$$g(x) = \frac{\beta}{\Gamma(\alpha/\beta)} \exp[-\alpha(x-\mu) - \exp\{-\beta(x-\mu)\}]$$

where Γ denotes the gamma function⁸, $\alpha(>0)$, $\beta(>0)$, and $\mu(-\infty < \mu < \infty)$ are three parameters (Coale and McNeil 1972). More precisely this was derived as the PDF of a random variable X such that

$$X = \left(\mu - \frac{\psi(\alpha/\beta)}{\beta} \right) + \sum_{j=1}^{\infty} \left(T_j - \frac{1}{\alpha + (j-1)\beta} \right)$$

where ψ denotes the digamma function, and T_j is a random variable that follows the exponential distribution with mean $1/\{\alpha + (j-1)\beta\}$ (Coale and McNeil, 1972).

Since model (1) given above is made up of infinite numbers of related exponential distributions, it can be viewed as a convolution of itself and some number of exponential distributions. According to equation of X above, the exponential distribution with the largest mean (smallest hazard) in the CM distribution has the parameter (hazard) α . Elimination of this distribution from the CM distribution makes another CM distribution with α substituted by $\alpha + \beta$. In general removal of the exponential distributions with the m -th largest means makes another CM distribution with α substituted by $\alpha + m\beta$. Specifically, letting $h_T(t; m)$ denote the PDF of the convolution of the exponential distributions with the m -th largest means and $g_X(x; m)$ be the

⁸ The gamma function is here defined as:

$$\Gamma(y) = \int_0^{\infty} u^{y-1} e^{-u} du.$$