

世紀中は人口増加が継続すると推計されている。

将来推計人口の変動は出生率の見通しに大きく依存するため、出生率の仮定設定が重要なものとなる。国連世界人口推計 2010 年版では出生率の中位仮定の算出のために確率モデルを採用した (Raftery et al. 2009、Alkema et al. 2010、Alkema et al. 2011)。詳細は次節にまとめるが、簡単に概観すると、過去の(合計)出生率のトレンドに基づき修正した確率要素の事前 (initial) 分布を仮定する。次にある国の過去の出生率トレンドと世界のその他の全ての国の傾向を加味して計算を行う。およそ 10 万本の将来の出生率の軌道を各国で推計し、その中の中央値を中位仮定値として設定している。出生率は長期的には人口置換水準に近づくように仮定している(低死亡率の国の出生率は 2.1 に近づくように水準を決定している)。その結果、中位仮定値における将来の出生率は出生力の水準によるグループ間で大きな差を生み出すことになる。

平均寿命は出生力 3 グループにおいてそれぞれ上昇すると仮定されている。2005 年-2010 年において、高出生力国で平均寿命が最も低く (56 年)、これは主に HIV/AIDS による。推計では将来的に他の主要な死因と同様に HIV/AIDS による死亡率も低下すると仮定しているため、高出生力国の平均寿命は 2045-2050 年で 69 年、2095-2100 年で 77 年になるとされる。中出生力国では 2005-2010 年で 68 年、2045-2050 年で 77 年、2095-2100 年で 82 年、低出生力国では 2005-2010 年で 74 年、2045-2050 年で 80 年、2095-2100 年で 86 年と推計されている。全体では平均寿命は 2005-2010 年で 68 年、2095-2100 年で 81 年と推計されている。

出生率の低下と寿命の伸長は人口高齢化を伴い、低出生力国で高齢化の進捗が最も早くなる。2010 年推計時点で低出生力国の人口の 11%が 65 歳以上人口であり、34%が 25 歳未満人口である。これが 2050 年までに、中位仮定値によれば 26%が 65 歳以上人口となり、24%が 25 歳未満人口となる。しかし、この仮定では低出生力国の出生率は増加すると仮定されているため、2100 年時点では、25 歳未満人口は 27%に増加し、65 歳以上人口も 28%に増加することに留意する必要がある。

3. 出生率の推計方法の概要

(Raftery et al. 2009、Alkema et al. 2010、Alkema et al. 2011)

ここでは、国連世界人口推計 2010 年版で採用された確率推計による出生率推計の方法について概要をまとめる。詳細は Raftery et al. (2009)、Alkema et al. (2010)、Alkema et al. (2011)において報告書ならびに論文が公表されており、統計ソフト R におけるパッケージ (bayesTFR_1.4-1、bayesDem_1.3-2) が公開されており、具体的な手法についても公開するとともに、利用者が異なる仮定を適用した推計結果を示すこともできる (例えば、低出生力国のみの傾向を反映した日本の出生率の推計など)。

国連世界人口推計 2010 年版では、確率人口推計の第一歩として、すべての国の合計出生率 (TFR) の推計に確率推計の手法を導入すると示されている (Raftery et al. 2009)。こ

のように、将来的には総人口の確率推計を射程に入れて手法の開発を行っている。

2010年版の出生率推計の概要を以下のようにまとめることができる。今回の出生率推計は、全ての国について、国別の TFR の推計についてベイズ推計モデルを用いて作成した。

このモデルは TFR の評価を 3つのフェーズに分解する。

- ・ 出生力転換前の高出生力 (pre-transition high fertility)
- ・ 出生力転換 (the fertility transition)
- ・ 出生力転換後の低出生力 (post-transition low fertility)

出生率の低下についてのモデルは国連人口部の決定論的推計手法を用いる。この手法は出生率が最終的に人口置換水準以下に低下することを仮定している。これは現在の TFR の水準に依存した 2段階ロジスティック関数の合計としての TFR の低下と誤差項 (random term) をモデル化するものである。ベイズ階層モデル (Bayesian hierarchical model) では、ある国の TFR の推移と全ての国のパターンの両方を用いて将来の TFR の推計を行う。これは国連が推定するマルコフ連鎖モンテカルロ法を用いた全ての国における過去の TFR の推移 (分布) から推定される。出生力転換後の低出生フェーズにおいては、TFR は人口置換水準付近を変動するとする。以上の手法の妥当性については、1980年からと1995年からの out-of-sample 推計によって評価を行う。

では、今回の世界人口推計に際して、出生率の推計に確率的手法を用いる理由はなぜであろうか。Alkema et al. (2011) では、以下のように説明している。

将来推計は政策担当者にとって政策決定のために重要であるのみならず、将来の不確実性、すなわち将来の結果が可能性としてどのような値をとるのかについて知ることも重要である。出生率は人口の規模と構造を決定する鍵となる。出生率低下は第一に人口の高齢化を決めることから高齢化の進度 (pace) も含めた将来の人口構造に重要な示唆を与える。合計出生率 (TFR) は人口推計における主要な要素の一つである。これは、女性が再生産年齢期間を通して生存し、その期間において各歳の年齢別出生率に従って子どもを産んだときの、女性一人当たりの平均的な子ども数になる。これまでの最も一般的な人口推計においては決定論的手法が採用され、スタンダードなコーホート要因法によるモデルが使用されてきた。将来の出生率も基本的には、決定論的に割り当てられてきた。国連人口部は出生率転換のさまざまなステージにある 196ヶ国の TFR の推計を行っており、出生率の推計は 2年ごとに修正している。決定論的な出生率の推計というのは、出生スケジュールに従って年齢別出生率の推計を行い中位仮定 (the medium variant)、最終的にコーホート要因法のもとで、死亡と国際人口移動とともに年齢別に推計が行われる。高位仮定は 0.5 加え、低位仮定は 0.5 減じることで設定している。この高位・低位の仮定は将来の出生率の水準の不確実性を評価するものではない。また、国単

位では更に不確実性が增大する。

このように 2008 年版までの出生率推計では、各国の低・中・高出生力国の全体のトレンドは表現できるものの、その平均からはずれた個々の国のトレンドは考慮できていなかった。そこで、2010 年版では、全体のトレンドをつかむこれまでの決定論的手法を用い、それを各国のトレンドを表現する手法として確率推計を導入している。具体的には、出生力転換によって出生率が低下している国には、系統的な変化をモデル化し、出生率低下の系統的低下の進度については、決定論的手法を基礎として関数をモデル化している。その低下パラメータの推定にはベイズ階層モデルを用いている。最終的に人口置換水準付近に到達した場合は時系列モデルを当てはめトレンドを推計し、長期的に人口置換水準付近で変動するように仮定している。各国がとりうる出生率の 3 フェーズに対応したモデル化を適用することによって、各国の過去のトレンドを再現、説明できるものとなっている。

3-1. 二段階ロジスティック関数による出生率減少関数の推定 (2008 年版モデル)

国連推計 2008 年版は、5 年間隔 5 歳階級別データを用いて 1950-2050 年まで出生率の推計を行っている。

推計手法を簡単にまとめると、c 国の TFR の 5 年後の低下の程度を示すパラメータの推計を行うことで将来の出生率を推計する。

$$f_{c,t+1} = f_{c,t} - d(\theta, f_{c,t}), \quad (1)$$

$f_{c,t}$ は c 国の 5 年期間 t における TFR

$d(\cdot, \cdot)$ は出生率転換モデルの減少関数パラメータ：現在の TFR 水準とパラメータ θ との関数として 5 年間の低下を示す。

この減少関数は二段階ロジスティック関数となっており、一段階のロジスティック関数は、急速に低下する高出生率から緩やかに低下する低出生率への減少分布を示す (図 1)。二段階のロジスティック関数は、人口転換が開始時の出生率低下が緩やかになる分布を示す。国連人口部では 6 パラメータから 3 つのパラメータを用いている。

$\theta \in \{\theta_{SS}, \theta_{FS}, \theta_{FF}\}$ S: Slow, F: Fast

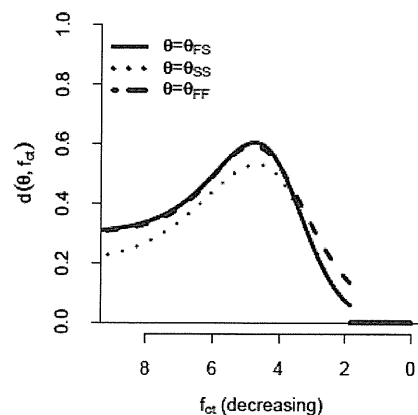


図 1 二段階ロジスティック関数 (Raftery et al. 2009)

人口置換水準以下では 1.85 で一定になるようにモデリングされている。このモデルは、

以下の2つの特性と2つの問題点がある。

特性は、

- (1) 決定論的モデルである。
- (2) 高出生率国が3つのパラメータから減少関数を選択。

問題点は、

- (1) 各国の変動を考慮していない。
- (2) 3つのパラメータは過去のトレンドも考慮していない。

2010年版では以上の2つの特性を考慮し2つの問題点を克服するモデルを適用する。

3-2. ベイズ階層モデル

2010年推計では、各国特有の出生率の確率推計モデルの作成にあたり、2008年版の国連モデルにおける出生率トレンドの3フェーズモデルを用いて基本モデルを設定している。

- (1) 出生力転換以前の高出生力：ベイズ階層モデルを用いない
- (2) 出生力転換：高出生力から人口置換水準またはそれ以下への低下
- (3) 出生力転換後の低出生力：人口置換水準以下からの回復、同水準近辺での変動

フェーズ2の出生力転換においては、国連の推計モデルを修正して推計し、出生力転換が進行している国においては系統的な低下を示すようにモデル化される。この低下モデルは変動を加えたランダムウォークモデルを適用し、以下のように定式化される。

$$f_{c,t+1} = f_{c,t} - d_{c,t} + \varepsilon_{c,t}, \text{ for } \tau_c \leq t < \lambda_c, \quad (2)$$

$$\varepsilon_{c,t} \sim N(0, \sigma_{c,t}^2), \text{ for } t \neq \tau_c, \quad (3)$$

$d_{c,t}$: drift 項 (出生力転換中の系統的低下をモデル化)

$\varepsilon_{c,t}$: ランダム変動 (系統的低下から生じる偏差)

τ_c : 出生率低下の始点 (

λ_c : 出生力転換後 (フェーズ3) の始点

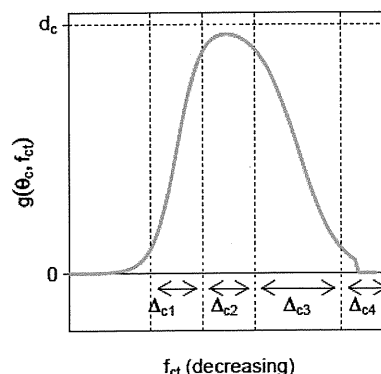


図2 修正二段階ロジスティック関数 (Raftery et al. 2009)

上記を二段階ロジスティック関数に組み込むと、

$$d(\theta_c, \lambda_c, \tau_c, f_{c,t}) = \begin{cases} g(\theta_c, f_{c,t}) & \text{for } \tau_c \leq t < \lambda_c \text{ and } f_{c,t} \geq 1; \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

$\lambda_c - 1$: 出生率低下の始点

$g(\theta_c, f_{c,t})$: 各国の二段階ロジスティック関数の固有パラメータ

$$\theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$$

$$\frac{-d_c}{1 + \exp\left(-\frac{2\ln(9)}{\Delta_{c1}}(f_{c,t} - \sum_s \Delta_{cs} + 0.5\Delta_{c1})\right)} + \frac{d_c}{1 + \exp\left(-\frac{2\ln(9)}{\Delta_{c3}}(f_{c,t} - \Delta_{c4} - 0.5\Delta_{c3})\right)}$$

修正ロジスティック関数を用いることで、各国の出生率低下モデルは固有の分布をもつことになる。その結果、図3のように2008年版では全体のトレンドから離れているような国の出生率低下モデルをその国特有のモデル推定結果となる。

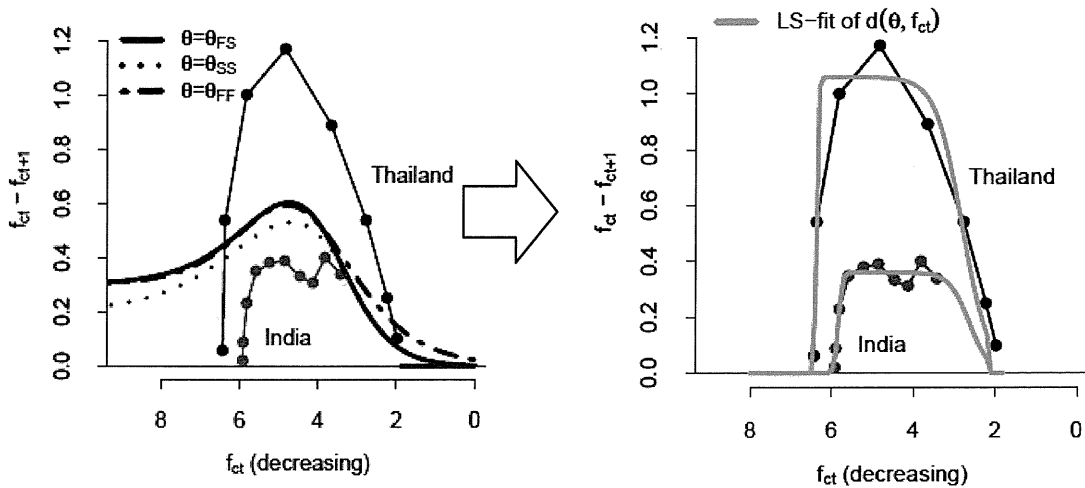


図3 2008年版二段階ロジスティックモデル(左)と修正二段階ロジスティックモデル(右) (例) インド、タイ (Raftery et al. 2009)

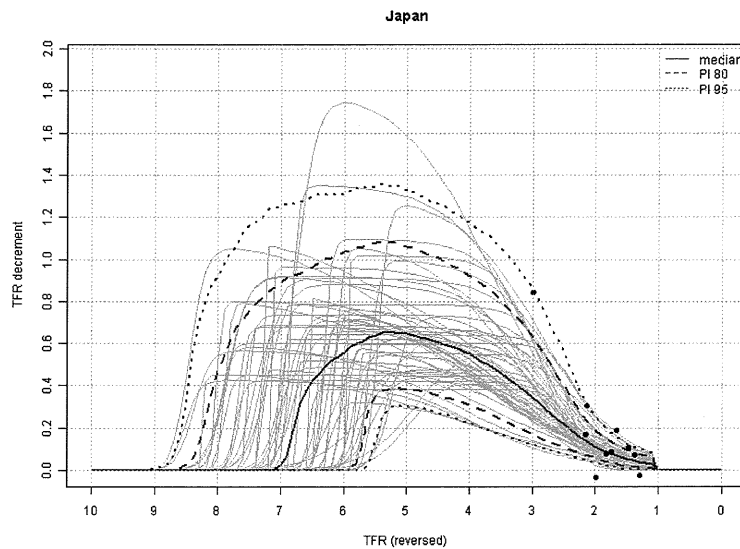


図4 修正ロジスティック関数によって推定される日本の出生率減少分布
 ・は実測値、実線は中央値、---は80%区間、...は95%区間
 (United Nations 2011b)

3-3. 自己回帰時系列 (AR(1)) モデル

フェーズ3の出生力転換後の低出生力については、長期的に人口置換水準近辺で出生率が変動するモデルを用いており、自己回帰時系列モデル (AR(1)モデル) $\mu=2.1$ である。

$$f_{c,t} \sim N(\mu + \rho(f_{c,t-1} - \mu), s^2) \text{ for } t > \lambda_c. \quad (5)$$

$$f_{c,t} = f_{c,t-1} + (1 - \rho)(\mu - f_{c,t-1}) + e_{c,t},$$

$$e_{c,t} \sim N(0, s^2),$$

ρ : 自己回帰パラメータ ($|\rho| < 1$ and s^2 : 誤差項の分散)

TFR が ρ に従って 2.1 から離れると、2.1 に向かって上昇・低下する。
AR(1)モデル $100(1 - \alpha)\%$ において、予測範囲は以下ようになる。

$$\left(2.1 - z_\alpha \frac{s}{\sqrt{1 - \rho^2}}, 2.1 + z_\alpha \frac{s}{\sqrt{1 - \rho^2}} \right), \quad (6)$$

$$z_\alpha : (1 - \frac{\alpha}{2}) \text{ 正規分布の四分位}$$

以上のような定式を用いてパラメータの推定を行う。データは国連が推定した 5 年間隔のデータセットを用いる。データセットは TFR の低下トレンドと予測範囲を全ての国について作成している。

t 辞典の国連推定値 u は TFR と同値であると仮定すると、以下ようになる。

$$u_{c,t} = f_{c,t}, \text{ for } t = 1, \dots, T_c, \quad (7)$$

T_c : c 国の観測期間数

3-4. パラメータ推定

出生力転換モデルのパラメータ推定には、観測期間を 2 フェーズと 3 フェーズに分ける必要ある。出生率の回復開始の定義は、TFR が 2 以下で 2 期間連続上昇が観測された時点としている。フェーズ 3 の c 国は以下のように定義される。

$$\begin{cases} f_{c,t} > f_{c,t-1}, \\ f_{c,t+1} > f_{c,t}, \\ f_{c,p} < 2 \text{ for } p = t-1, t, t+1. \end{cases}$$

フェーズ 2 と 3 を分けることができない場合は、両方のパラメータが推定される。パラメータは国連が作成した低下パラメータについての事前分布からマルコフ連鎖モンテカルロ法 (MCMC) によって、出生力転換モデルのそれぞれのパラメータの事後分布を得る。サンプリング方法は、Gibbs sampling、Metropolis-Hastings、slice sampling step の混合アルゴリズムとなる。

フェーズ 3 の AR(1)パラメータ ρ と s は最尤推定法によって推定している。出生率の反転が観測された国は 20 개국 (シンガポール、ブルガリア、チェコ、ロシア、チャネル諸島、デンマーク、エストニア、フィンランド、ラトビア、ノルウェー、スウェーデン、イギリス、イタリア、スペイン、ベルギー、フランス、ドイツ、ルクセンブルク、オランダ、アメリカ合衆国) となり、最尤推定値 $\rho = 0.906$ 、残差の標準偏差 0.09 となっている (人口転換後出生率の回復がみられる 8 か国の 4 期間を算出した結果)。

TFR の推計には、MCMC 法によって得られたパラメータから得られる事後分布のサンプルを用いる。出生力転換が完全に終わっていない国で、TFR が $\Delta c4$ 以下にまで低下しているとフェーズ 3 は開始しない (図 2 参照)。TFR の観測値が 2 時点で回復するとフェーズ 3 の AR(1)モデルの当てはめが始まる。出生率回復過程 λc の始点は $t > Tc$ として、以下のように定義される。

$$\begin{cases} f_{c,t} > f_{c,t-1}, \\ f_{c,t-1} < \Delta c4. \end{cases}$$

$\Delta c4$ の水準で、出生率の回復が始まった時点ではフェーズ 2 のモデルが適用される。出生力転換が $\lambda c-1$ 時点において収束した時点でフェーズ 3 のモデルが適用され、AR(1)モデルによる出生率の推計がなされる。AR(1)モデルにおける変動の標準偏差は限られた期間数によって推定されるため、0.203 とばらつきは大きい。それぞれの国の将来の TFR の軌道は約 10 万本の軌道として出力される。その中の中央値を中位推計値として採用する。推計値の範囲は 80%、95%信頼区間、予測範囲は 10%tile から 90%tile、2.5%tile、97.5%tile の結果が示されている。

3-5. TFR の推計結果と妥当性評価

TFR の推計結果は国・地域ごとに参照することができる。将来の不確実性の幅は、出生率グループ (高出生力・中出生力・低出生力) によって異なり、高出生力国は TFR の値が高い国ほど 80%信頼区間における 2045-50 年と 2095-2100 年での変動が大きくなり、低出生力国では TFR の値が小さい国ほど変動が大きくなる。最も変動が少ないのは TFR が 2 ~ 3 の水準にある国となる。日本についての結果は図 6 に示している。2006 年以降の出生率の回復傾向を反映して、今後は出生率が長期的に人口置換水準に向かって収束するような出生率の軌道が得られている。

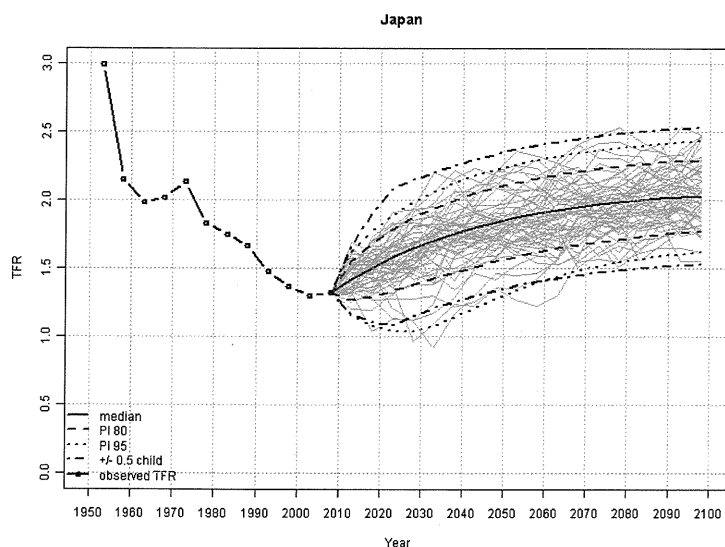


図 6 日本の出生率の推計結果 (United Nations 2011c)

モデルの妥当性評価には、out-of-sampling 推計法を用いている。この妥当性の検証は、1975-80年（5年ごと6期間について推計）と1990-95年（5年ごと3期間について推計）を始点としてベイズ推計モデルによる出生率の推計を行い、それぞれ実績値との比較を行い評価するものである。モデルの妥当性の基準は、80%信頼区間で10%以内の変動、95%信頼区間で2.5%以内の変動としている。以上の方法で評価を行った結果、予測範囲の信頼区間はうまく推定できているが、推計期間が長期になるほどTFRはやや過大に推計される性質があることが示された。

日本についての out-of-sampling 結果では（図7）、1975-80年を始点としたケースでは、2時点においてはほぼ推計されているが、それ以降は過大なものとなっている。一方、1990-95年をしたケースではやや過大であるが、概ね実測値に近い推計結果となっている。

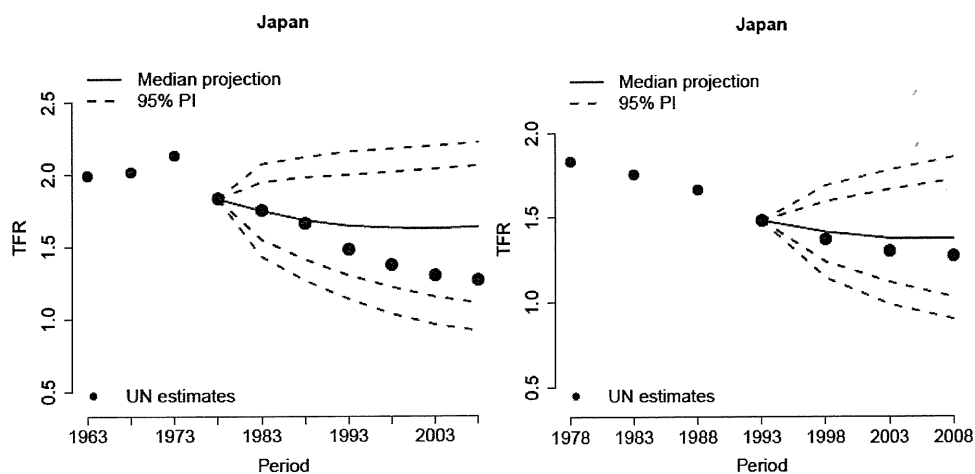


図7 out-of-sampling による日本の出生率推計モデルの妥当性 (Raftery et al. 2009)

3-6. おわりに

ここまで、国連経済社会局人口部が公表した国連世界人口推計2010年版の概要と新たに導入された出生率中位仮定設定のための確率推計についてその概要をまとめた。確率推計による出生率の推計結果は、出生率を取りうる信頼区間の設定には大いに役立つものの、中央値と実績値との比較では、やや過大に推計されがちな結果となっている。これは、第一に全世界の出生率の推移を考慮したモデルを用いていること、第二に出生率が超長期的には人口置換水準近辺に収束するためである。

参考文献

- Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F. and T. Buettner (2010) “Probabilistic Projections of the Total Fertility Rate for All Countries”, *Working Paper* No. 97, Center for Statistics and the Social Sciences, University of Washington.
- Alkema L., A. E. Raftery, P. Garland, S. J. Clark, F. Pelletier, T. Buettner, and G. K. Heilig (2011) “Probabilistic Projections of the Total Fertility Rate for All Countries”, *Demography* 48, pp.815-839.
- Raftery E. Adrian et al., (2009) “White Paper: Probabilistic Projections of the Total Fertility Rate for All Countries for the 2010 World Population Prospects”, from “Expert Group Meeting on Recent and Future Trends in Fertility” UN Nov. 12, 2009.
- United Nations, Department of Economic and Social Affairs, Population Division (2011a) “World Population Prospects: The 2010 Revision, Press Release (3 May 2011)”.
http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Press_Release.pdf
- United Nations, Department of Economic and Social Affairs, Population Division (2011b) “Projections of Total Fertility based on Bayesian Hierarchical Model (BHM)- Decline Curves (based on Double-Logistic Function) from BHM, Median, 80% and 95% prediction intervals-”.
http://esa.un.org/unpd/wpp/fertility_figures/interactive-figures_DL-functions.htm
- United Nations, Department of Economic and Social Affairs, Population Division (2011c) “Projections of Total Fertility based on Bayesian Hierarchical Model (BHM)- Projections of Total Fertility: Median, 80% / 95% prediction intervals and high / low WPP fertility variant-”.
http://esa.un.org/unpd/wpp/fertility_figures/interactive-figures_TF-trajectories.htm
- 石川晃、佐々井司、別府志海 (2011) 「国連世界人口推計 2010 年版の概要」、『人口問題研究』, 第 67 卷, 第 3 号, pp.13-50.

資料 1 bayesTFR パッケージ (バージョン 1.4-1) の主要関数の概説

作者 Hana Sevcikova, Leontine Alkema, Adrian Raftery

条件 R (2.9.0 以上), mvtnorm, coda

弱条件 rworldmap, fields, snowFT, googleVis

記述 ベイズ階層モデルを用いて, 世界の全ての国の合計出生率を確率推計によって作成する.

URL <http://bayespop.csss.washington.edu>

bayesTFR-package ベイズ出生率推計

ベイズ階層モデルを用いて, 世界の全ての国の合計出生率を確率推計によって作成する関数群.

推計は Alkema らによる手法による. 国連によって作成された時系列 (historical) データを使い, 全世界の全ての国の合計出生率の事後分布 (posterior distribution) を同時にシミュレートする.

主要な関数は

- **run.tfr.mcmc** : マルコフ連鎖モンテカルロ法 (MCMC) 可能な限り並列に走らせる関数. その結果, MCMC パラメータの事後サンプルを得る事ができる. 現在あるシミュレーションは **continue.tfr.mcmc** を用いることによって再計算することができる.
- **tfr.predict** : 事後パラメータサンプルを用いることによって, 全ての国の合計出生率の推計の軌道を得る事ができる.
- **run.tfr.mcmc.extra** : ベイズ階層モデルに含まれていない国など特別な国または地域について MCMC を実行する. 集合体にも適用することができる.
- **tfr.predict.extra** : 特別な国または結合された地域の予測値を生成する.

上記の関数の順番は, このパッケージを利用するときのおおまかな作業順となっている. 1) MCMC シミュレーションを実行, 2) 予測値を生成, 3) 結果の分析 (以下の関数を使用). 予測に使用していない国や地域の場合は, **extra** 関数を用いて推計ができる.

結果を分析する関数もパッケージには含まれている.

- **tfr.trajectories.plot** : ある国の事後の軌道 (中央値と確率的な範囲) をプロットする.
- **tfr.trajectories.table** : ある国の事後の軌道を表形式で出力する.
- **tfr.map** : 推計期間における TFR の世界地図を表示する.
- **DLcurve.plot** : シミュレーションに使用する二段階ロジスティック関数の事後カーブ (中央値と確率的な範囲) を示す.
- **tfr.partraces.plot** と **tfr.partraces.cs.plot** : 国独自のパラメータと国別のパラメータの MCMC による軌道をそれぞれ表示する.

- **tfr.pardensity.plot** と **tfr.pardensity.cs.plot** : 国独自のパラメータと国別のパラメータの MCMC による密度をそれぞれ表示する.
- **summary.bayesTFR.mcmc.set** : MCMC 結果についての関数の要約.
- **summary.bayesTFR.prediction** : 予測値の結果についての関数の要約.

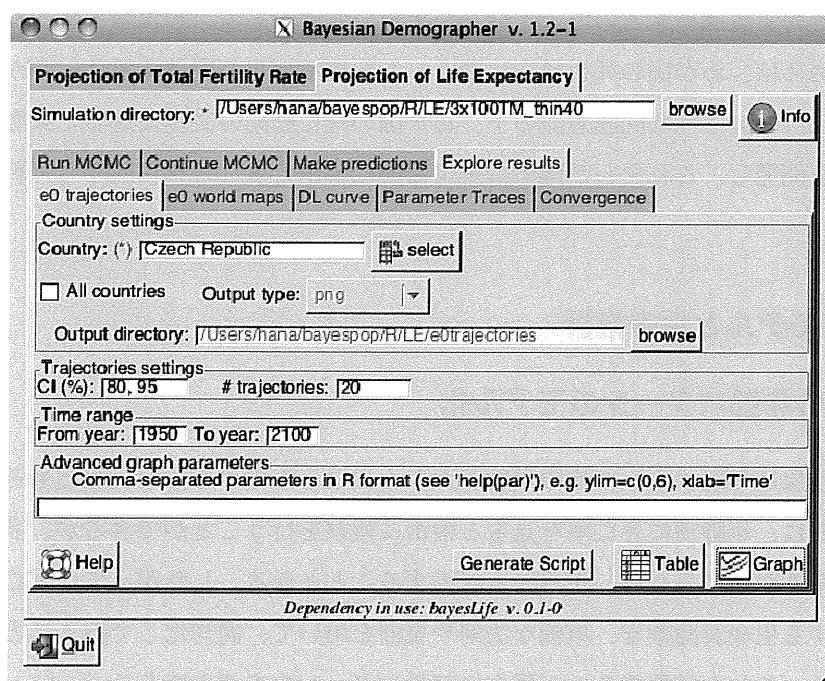
MCMC の診断については, **coda.mcmc.list** 関数は「mcmc.list」のタイプを作成し, これは coda パッケージを使用する. また, **tfr.diagnose** 関数は, coda パッケージにある Raftery の診断法を用いて MCMC を分析し, 収斂しなかったパラメータについての情報を与える. 多数の連鎖をもつケースについては, **tfr.gelman.diag** が Gelman の診断に基づき, **tfr.plot.gelman** を用いてプロットすることによって確認することができる.

現在のシミュレーション結果は **get.tfr.mcmc** 関数, 現在の予測値は **get.tfr.prediction** 関数を用いることによって得ることができる. 現在の収斂についての診断についても, **get.tfr.convergence** や **get.tfr.convergence.all** 関数によって得ることができる.

使用例)

```
# ディレクトリ内の ex-data のような結果を作成するコマンド
m <- run.tfr.mcmc(nr.chains=3, iter=50, seed=1, verbose=TRUE)
pred <- run.predict(m, burnin=30, verbose=TRUE)
summary(pred, country='Ghana')
```

資料 2 bayesDem_1.2-1 パッケージの GUI



(参照) http://bayespop.csss.washington.edu/data/bayesLife/#using_bayesDem

9 死亡率曲線の自由な方向への変化を表現する数理モデルとわが国の将来生命表への応用

石井 太

はじめに

本研究事業では、外国人人口受入れによる将来人口の変化に関するシミュレーションの実行を一つの目的としている。このような定量的シミュレーションの実行には、将来の死亡率の見通しを定量的に評価することが必要である。わが国ではこれまで、外国人人口受入れに関して比較的保守的な政策を採ってきたことから、日本に受け入れた外国人が将来的にどのような死亡水準となるかを推計するために直接用いることのできるデータは限られている。しかしながら、入国した外国人は、日本に居住している間、日本人と同様の生活環境で過ごすこと、同時にわが国の高い公衆衛生・保健医療水準の恩恵を享受できることなどを考えると、長期的には日本人の死亡水準と一定の連動をしながら推移するものと考えられる。したがって、現状の傾向が続くとした場合の死亡動向の分析とモデリング、またそれに基づく将来見通しに関する研究を行っておくことは、このようなシミュレーションのための前提として重要である。

ところで、わが国の平均寿命は、国際的に見てトップクラスの水準にありつつも、さらに伸長を続けているという特徴を有している。このような、先進諸国の中でも独特の様相を呈している死亡率の将来推計にあたっては、これまで諸外国等で用いられてきている死亡モデルをそのまま適用するのでは十分とさえいえない側面がある。したがって、シミュレーションのためには、わが国の死亡状況に適合した数理モデルを開発することが重要となる。本研究は、このような問題意識の下、わが国の死亡動向の表現に適した柔軟性の高い数理モデルを開発し、これをわが国の将来生命表に応用することを目的として行ったものである。

1 死亡率モデルと将来推計

1.1 死亡の年齢パターンのモデル化

人口動態事象は年齢によってその頻度が大きく異なるものがあるが、死亡もその代表例である。一般に、年齢別に死亡率を見ると精密に観察を行うことができるが、多くの率を取り扱う必要性が生じるため、複雑さも増大することとなる。したがって、死亡率などの死亡を表現する生命表関数を、単純な法則や規則を用いて、簡約化して表すことができると便利であり、これを実現するのが死亡の年齢パターンのモデルである。

死亡の年齢パターンをモデル化すると、直接測定可能なデータが限られていたりデータの精度が低い場合に、モデルによる間接的な測定を行うことや、標準的なパターンから外

れた特異値をモデルを用いて特定することなどが可能となる。また、複数の異なる死亡状況を比較する場合、モデルに含まれるパラメータの違いを解釈することによって、両者の違いの要因を分析することも可能となる。

死亡の年齢パターンのモデルには、大きく分けて、数学的関数によるモデル、数表によるモデル、リレーショナルモデルの3種類がある。数学的関数によるモデルとは、年齢の数学的関数によって生命表関数を表現するものであり、死亡法則 (law of mortality) とも言われる。代表的な例として、死力を年齢の指数関数で表す Gompertz モデルがある (Gompertz 1825)。一方、数表によるモデルは、経験に基づくいくつかの数表によって生命表関数を表現するものである。Coale-Demeny のモデル生命表 (Coale and Demeny 1983) が最もよく知られている例であり、生命表の形状 (シェイプ) を4種類 (North, South, East and West) に分け、それぞれに25のレベルから成るモデル的な生命表を用意して様々な生命表を表現するモデルである。数学的関数によるモデルは、生命表関数を数少ないパラメータで表現でき、豊富な情報を簡約化して表現できるという点で優れているが、一方で、実際の死亡の年齢パターンは複雑であり、必ずしも既知の数学的関数で表されるとは限らない。また、これを改善するために複数の数学的関数を組み合わせていくと、多くのパラメータが必要となつてかえって複雑性を増してしまうという問題点がある。これに対し、数表によるモデルは、実際に経験された年齢パターンから作成されることから、このような数学的関数によるモデルのような制約はない。しかしながら、この方法では、それぞれのレベルや形状の種類に対応した数表を用意しなければならないため、レベルや形状を細分化していくと、膨大な数の数表が必要となり、やはり複雑性が増大してしまうという問題点がある。

リレーショナルモデルは、この二つのモデルの長所を取り入れたモデルともいうことができ、経験に基づく (数表による) 標準的な年齢パターンと、そこからの変化に関するパラメータを組み合わせて死亡パターンを表現するものである。リレーショナルモデルの代表例として、Brass (1971) により開発されたブラス・ロジットシステムが挙げられる。ブラス・ロジットシステムは、 l_x をロジット変換した関数 $Y_x = \log\left(\frac{l_x}{1-l_x}\right)$ を考え、ある基準となるパターン Y_x^s を用いて、任意の Y_x^a が、 $Y_x^a = \alpha + \beta Y_x^s$ と表されるとするモデルである。ここで、 α がレベル、 β が形状を表すパラメータとなっている。

1.2 死亡の年齢・時系列パターンのモデル化

1.1 節においては、単に年齢パターンを考えてきたが、死亡率の将来推計を行う観点から、死亡の年齢・時系列パターンのモデル化について考えよう。これは、例えば、生命表関数として対数死亡率 $\log m_x$ をモデリングする場合、さらに時刻 (t) を新たな次元として追加した $\log m_{x,t}$ をモデリングする問題を考えることになる。また、時系列パターンについては、数学的関数で表現されるなど、補外による将来投影が可能となるものを考察の対象とする。

死亡の年齢・時系列パターンモデル化に対しても、いくつかのアプローチが可能である。最も単純なものとしては、各 x を固定し、 $\log m_{x,t}$ を t の関数と考える数学的関数を当てはめるモデリングが考えられる。しかしながら、このためには年齢の数だけ異なる数学的関数を用意する必要が生じるため煩雑なモデルとなってしまう。そこで、1.1 節で考えた年齢パターンモデルを組み合わせることが考えられる。例えば、年齢パターンについて数学的関数によるモデルを用い、このパラメータの時系列変化にさらに数学的関数を当てはめれば、年齢・時系列パターン全体を数学的関数を用いてモデリングすることができる。また、モデル生命表との組合せとして、生命表の形状は固定し、レベルを表す指標（平均寿命など）に数学的関数を当てはめるモデリングも可能である。

リレーショナルモデルについても、数学的関数によるモデル同様、パラメータに何らかの数学的関数を当てはめることによって年齢・時系列パターンモデル化が可能である。特に、リレーショナルモデルは、標準的な年齢パターンとそこからの変化によってモデリングされていることから、後者に相当する部分は死亡年齢パターンの時系列変化分をモデリングしたことに相当する。例えば、リー・カーター・モデルは以下のような式で表されるリレーショナル・モデルの一種であり、現在、国際的にも死亡率の将来推計の標準的な方法として使われている (Lee and Carter 1992)。

$$\log m_{x,t} = a_x + k_t b_x + \epsilon_{x,t}$$

ここで、

$\log m_{x,t}$: 対数死亡率

a_x : 対数死亡率の標準的な年齢パターン

k_t : 死亡水準（死亡指数）

b_x : k_t が変化する時の年齢別死亡率の変化

$\epsilon_{x,t}$: 平均 0 の残差項

である。

k_t が数学的関数でモデル化されているとすれば、この式は、 $\log m_{x,t}$ が、標準的な年齢パターン a_x と、そこからの時間変化 $k_t b_x$ に分解してモデル化されたことを示している。ここで、標準的なパターンは一定であることから、リレーショナルモデルを用いた対数死亡率のモデリングとは、「対数死亡率の時間変化のモデリング」に相当するのである。

このことについて、具体的にリー・カーター・モデルで見よう。上の式を誤差項を無視して t で偏微分することにより、

$$\frac{\partial \log m_{x,t}}{\partial t} = \frac{\partial k_t}{\partial t} b_x$$

を得る。これは、対数死亡率の時間変化が、全体の強度は $\frac{\partial k_t}{\partial t}$ により異なるものの、その年齢分布は b_x で一定であることを示している。この符号を変えたものは、一般に死亡率改善率と呼ばれる量であることから、リー・カーター・モデルは死亡率改善率の年齢分布が一定であるモデルであるということもできる。

ところで、死亡率改善率が示す「時間変化」とは、それぞれの x について対数死亡率がどのくらい低下するかという捉え方である。一方、Bongaarts (2005) は先進諸国の死亡率にロジスティック曲線を当てはめ、そのパラメータの観察を通じて、先進諸国の成人死亡率改善を「低下」として捉えるより、死亡率曲線の高齢側への「シフト」として捉えるべきであるとして、これを表現するシフティング・ロジスティック・モデルを提案した。また、わが国の高齢死亡率についても、「低下」より「シフト」として捉える方が効率的であることが石井 (2006)、石井 (2008) 等の先行研究において示されている。すなわち、わが国の高齢死亡率改善は、リー・カーター・モデルのように、 x を固定して対数死亡率が低下する方向に対数死亡率曲線が変化していくと考えるよりも、対数死亡率値を固定してこれに対応する年齢 x が高齢の方に変化していくという改善方向で捉えた方が効果的ということである。このように、「対数死亡率の時間変化のモデリング」において、その改善方向の新たな捉え方というのが必要となるのである。

そして、これをより一般化すると、ある時刻におけるある年齢の対数死亡率が、年齢も対数死亡率も異なる新たな方向へと変化していくという状況を考えることも可能である。そこで、このような死亡率曲線の自由な方向への変化を数理モデルによって表現するためのフレームワークについて整理し、さらに、これに対応した死亡率モデルの表現について考察しよう。

2 死亡率モデリングのフレームワーク

2.1 対数死亡率曲面、対数死亡率とその逆関数について

死亡率曲線の自由な方向への変化を表現する数理モデルの構築のため、対数死亡率曲面を用いた死亡率モデリングのフレームワークが必要となる。以下、これに関して必要な概念を定義する。

$X = [0, +\infty)$ を年齢、 $T = (-\infty, +\infty)$ を時間を表す変数とする。以下の死亡モデリングの議論においては、年齢 $x \in X$ 、時刻 $t \in T$ における死力関数 $\mu_{x,t}$ を用いる。また、

$$y = \lambda_{x,t} = \log \mu_{x,t}$$

で死力の対数を表す。ここで、 $y \in Y = (-\infty, +\infty)$ は対数死力関数の値を表す。このとき、 S を、

$$S = \{(x, t, y) | y = \lambda_{x,t}\}$$

とすると、これにより、 \mathbb{R}^3 における曲面が定義される。これを、対数死亡率曲面 (the log mortality surface) と呼ぶ。これは、対数死亡率曲面の一般的な表示方法である。この表示方法において、 $y = \lambda_{x,t}$ は、 \mathbb{R}^3 の中で、 X - T 平面からの高さとして捉えることができる。

以下、 $\lambda_{x,t}$ は、 $X_0 \times T_0 = [0, \omega] \times [t_0, t_1] \subset X \times T$ 上で定義された、 x と t に関する滑らかで連続な関数であると仮定する。ただし、 $\omega < +\infty$ は死亡率モデルを考える上での

ある有限な年齢の最大値とする。

成人死亡率のモデリングにあたっては、さらに、 $\lambda_{x,t}$ は各 t について、 $x > x_0(t)$ の範囲で x について狭義単調増加であると仮定してよい。ここで、 $x_0(t)$ は各 t について、 $\lambda_{x,t}$ が狭義単調増加関数となる x の下限値を示す。このとき、各 t に対して、関数 $\lambda_t(x)$

$$\lambda_t : \tilde{X}_t \rightarrow Y, \quad \lambda_t(x) \stackrel{\text{def}}{=} \lambda_{x,t}$$

は x について単射である。ただし、 $\tilde{X}_t = [x_0(t), \omega]$ である。 $\tilde{Y}_t = \lambda_t(\tilde{X}_t)$ とすると、関数 $\lambda_t(x) : \tilde{X}_t \rightarrow \tilde{Y}_t$ は各 t について、 \tilde{Y}_t 上で定義される逆関数 $\nu_t(y) : \tilde{Y}_t \rightarrow \tilde{X}_t$ を持つ。

そこで、 Y_0 を以下のように定義する*1。

$$Y_0 \stackrel{\text{def}}{=} [y_0, y_1] \quad \text{ただし} \quad y_0 = \sup_{t \in T_0} \min \tilde{Y}_t, \quad y_1 = \inf_{t \in T_0} \max \tilde{Y}_t,$$

このとき、 $\nu_{y,t} : Y_0 \times T_0 \rightarrow X_0$ を

$$\nu_{y,t} \stackrel{\text{def}}{=} \nu_t(y)$$

により定義する。 $\nu_{y,t}$ は、時刻 t における対数死力関数の値が y となるような年齢 x を示している。

さらに、死亡モデルの微分型表示において用いる、(1) $\rho_{x,t}$: 死力改善率と (2) $\tau_{y,t}$: 年齢増加率を、以下の式によって定義する。

$$\rho_{x,t} \stackrel{\text{def}}{=} -\frac{\partial \lambda_{x,t}}{\partial t} = -\frac{\partial \log \mu_{x,t}}{\partial t}$$

$$\tau_{y,t} \stackrel{\text{def}}{=} \frac{\partial \nu_{y,t}}{\partial t}$$

2.2 対数死亡曲面上の接ベクトル場を利用したモデル構築

1.2 節で述べた通り、リー・カーター・モデルのような低下型モデルは、年齢を固定した場合の対数死亡率の時間的変化を記述することによって死亡率の変化を表現する。一方、シフト型モデルは、対数死亡率値を固定した場合の年齢の時間的変化を記述するモデルと考えることができる。これらの事実を、対数死亡率曲面 S 上の接ベクトルを用いて表現しよう。

対数死亡率曲面 S 上で、年齢 x_0 を固定して移動した時に描かれる曲線 C_{x_0} と、対数死亡率値 y_0 を固定して移動したときに描かれる曲線 C_{y_0} を考える。このとき、 C_{x_0} と C_{y_0} はそれぞれ、

$$(x_0, t, \lambda_{x_0,t})$$

$$(\nu_{y_0,t}, t, y_0)$$

*1 $x_0(t)$ は必ずしも連続とは限らないため、 $\{\min \tilde{Y}_t | t \in T_0\}$ には最大値がない可能性もある。

とパラメータ表示されるから、これを t で微分して得られる以下の2つのベクトル

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

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

は、 S の接ベクトルとなる。

図1 対数死亡率曲面と接ベクトルの例

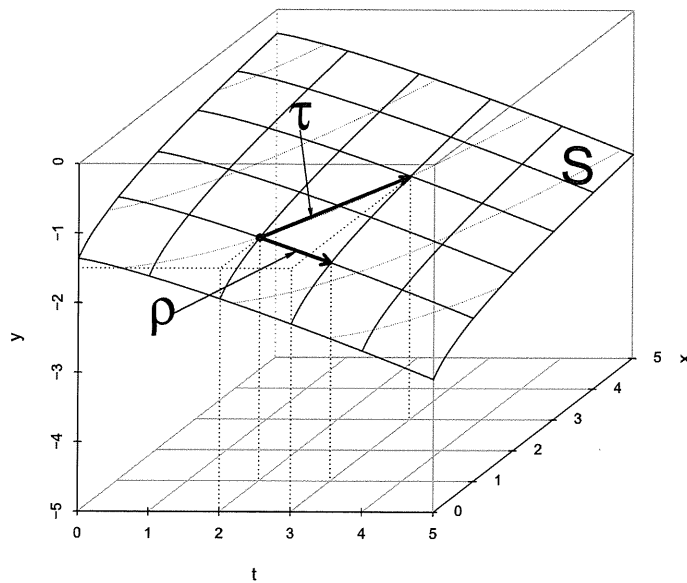


図1は、対数死亡率曲面 S 及びそれ上の接ベクトルの例を模式的に示したものである。対数死亡率曲面上の点 $(x, t, y) = (1, 2, -1.5)$ において、年齢を固定した方向への接ベクトルが ρ 、対数死亡率値を固定した方向への接ベクトルが τ である。

このような接ベクトルは、 S の任意の点で考えることができるから、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る。このように、対数死亡率の時間変化は接ベクトル場に対応させることができるが、逆に、各時刻・各年齢における死亡率の改善方向をベクトルで指定することによって、これを接ベクトル場として持つような対数死亡率曲面を構成することができる。したがって、各時刻・各年齢における死亡率改善方向をモデル化するという、接ベクトル場を利用した死亡モデルの新たな表現が得られる。

例えば、

$$\rho(x_0, t_0, y_0) = (0, 1, k'_t b_x)$$

とすれば、これがLCモデルを表現する接ベクトル場ということになる。

この死亡モデルの新たな表現を用いれば、各時刻・各年齢における改善方向をわが国の死亡状況に適合するようにモデリングできれば、好ましいモデルが構築できることになる。そこで、このようなアプローチに従って死亡率のモデリングを考えることとしよう。

2.3 シフト型モデルとしての線形差分 (LD) モデル

国立社会保障・人口問題研究所の「日本の将来推計人口（平成 18 年 12 月推計）」（国立社会保障・人口問題研究所 2007）の将来生命表作成にあたっては、現在国際的に標準的な方法とされるタリー・カーター・モデルを採用しつつ、これに対して世界の最高水準の平均寿命を示すわが国の死亡動向の特徴に適合させるため、新たな機構を加えた「年齢シフトモデル」により死亡率の投影を行った（石井 2008）。具体的には、過去の死亡率曲線にロジスティック曲線をあてはめて、その年齢シフト量と勾配に関するパラメータを推定し、これによる高齢死亡率の年齢シフトを考慮した上でリー・カーター・モデルを適用することによって、死亡率改善の著しいわが国の死亡状況に適合させるものである。

このようなモデルがわが国の死亡動向によく適合するのには、日本の近年の高齢死亡率改善が、死亡率曲線の垂直方向への低下よりも、水平方向にシフトするという見方がより実態に合っているということが背景にあった。そこで、対数死亡曲面上の接ベクトル場を利用したモデリングに応用しやすく、また、効率的に高齢死亡率を表現可能な「シフト」型モデルである線形差分 (LD) モデルを導入する。

まず、LD モデルが高齢死亡率を効率的に表現可能であるという理論的背景について述べる。高齢部の死亡率を表すパラメトリックなモデルに関してはこれまで多くのものが提案されている。このようなものの中から、Thatcher et al. (1998) は、13 カ国の 80～120 歳の高齢死亡率データに対し、6 種類のパラメトリックモデルを当てはめ、モデル間の比較検討を行っている。ここで比較対象となったモデルは、

1. Gompertz: $\mu_x = ae^{bx}$
2. Weibull: $\mu_x = ax^b$
3. Helligman & Pollard: $q_x = \frac{ae^{bx}}{1+ae^{bx}}$
4. Quadratic: $\ln(\mu_x) = a + bx + cx^2 \quad (c < 0)$
5. Logistic: $\mu_x = \frac{ae^{bx}}{1+ae^{bx}} + c$
6. Kannisto: $\mu_x = \frac{ae^{bx}}{1+ae^{bx}}$

であるが、この中で、もっともパフォーマンスが高かったものは、Logistic と Kannisto のロジスティックモデルであり、特に、Kannisto は、2 パラメータロジスティックモデルとも呼ばれ、2 つだけのパラメータで高齢死亡率をよく表現できるという利点があるとされている。また、このモデルは、Human Mortality Database の高齢部分のあてはめにも用いられており、先進諸国の高齢死亡率の表現に有効であることがわかる。したがって、このモデルが時系列的に変化する時に持つ特性を抽出して、死亡率改善方向に翻訳するこ

とができれば、接ベクトル場を利用しつつ、高齢死亡率を表現する上で好ましい性質を備えたモデルの構築が可能となることになる。

以下の命題は、2パラメータロジスティックモデルの $\tau_{y,t}$: 年齢増加率が持つ特性であり、LDモデルの考え方の基礎を示すものである。

命題 1. 2パラメータロジスティックモデル

$$y = \lambda_{x,t} = \log \frac{\alpha_t \exp(\beta_t x)}{1 + \alpha_t \exp(\beta_t x)} = \log \alpha_t + \beta_t x - \log(1 + \alpha_t \exp(\beta_t x))$$

において、 $\tau_{y,t}$ は各 t において x の線形関数である。すなわち、

$$\tau_{y,t} = f'_t + g'_t x$$

が成立する。

Proof.

$$e^y = \frac{\alpha_t \exp(\beta_t x)}{1 + \alpha_t \exp(\beta_t x)}$$

$$\Leftrightarrow \alpha_t \exp(\beta_t x) = \frac{e^y}{1 - e^y}$$

y を固定したまま両辺を t で微分することにより、

$$\alpha'_t \exp(\beta_t x) + \alpha_t \exp(\beta_t x) \left(\beta'_t x + \beta_t \frac{\partial x}{\partial t} \right) = 0$$

$$\Leftrightarrow \frac{\partial x}{\partial t} = -\frac{\alpha'_t}{\alpha_t \beta_t} - \frac{\beta'_t}{\beta_t} x$$

□

したがって、 $\tau_{y,t}$ が x の線形関数となる場合、2パラメータロジスティックモデルの年齢増加率が持つのと同じ性質を持つことになる。そこで、年齢増加率が実際にこの条件を満たすモデルとしてLDモデルを定義する。すなわち、LDモデルは、年齢増加率が以下のように表されるモデルである。

$$\tau_{y,t} = f'_t + g'_t x$$

これは微分型表示であり、両辺を t で積分することにより、以下の通常型表示を得る。

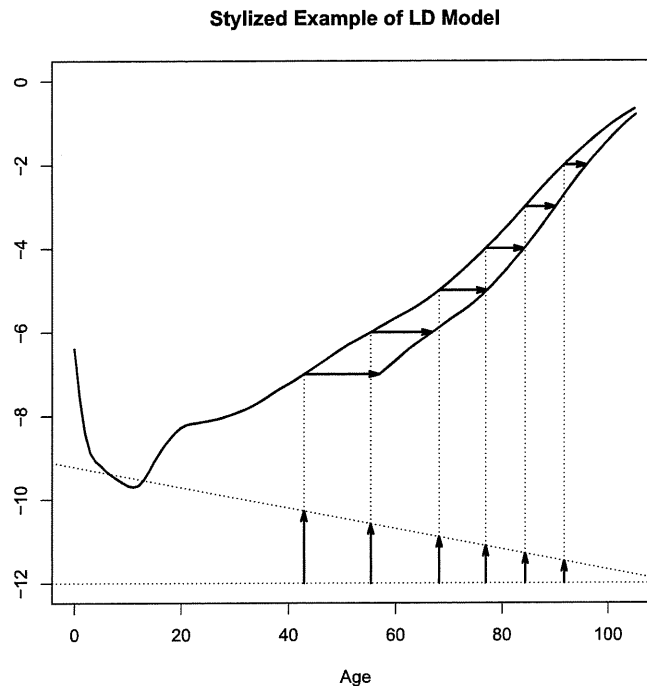
$$\nu_{y,t} = f_t + g_t x + a_y$$

ここで、 a_y は対数死亡率の逆関数に関する標準的なパターンを表す。

図2は、LDモデルの考え方を模式的に表したものである。図は、左側にある対数死亡率曲線の成人死亡率部分が、死亡率改善により右側へシフトした状況を示している。ここで、水平方向右向きの矢印が各年齢におけるシフト量、すなわち、年齢増加率を表す。この矢印を反時計回りに90度回転させ、始点を揃えてグラフの下側に配置した。矢印の終

点が一直線上に並んでいるが、これが年齢増加率が年齢の線形関数になることを意味している。LD モデルとは、この矢印の終点が常に直線上に並ぶモデルであるということができる。

図2 LD モデルの模式的例



LD モデルは、成人死亡率のみを表現するモデルであり、全年齢の死亡率を必要とする将来生命表の作成に単独で用いることができない。一方、LC モデルは全年齢の死亡率をモデリングすることが可能である。そこで、若年層に対しては LC モデルの改善方向を持ちつつ、高齢層に対しては LD モデルの改善方向を持つようなベクトルを構成し、これを接ベクトル場として持つような対数死亡率曲面として全年齢死亡率モデルを構築することが考えられる。次節以降において、このような方法を実際のわが国の死亡率に適用し、将来生命表に応用することについて考察する。

3 わが国の死亡率への適用

3.1 実績値によるモデル評価・構築

本研究において基礎とする死亡率データは、国立社会保障・人口問題研究所「わが国の長寿化の要因と社会・経済に与える影響に関する人口学的研究」プロジェクトにおいて作成された「日本版死亡データベース」とし、昭和 45(1970)～平成 22(2010) 年の死亡率を用いた。「日本版死亡データベース」とは、国際的な死亡データベースである Human Mortality Database の方法論を基礎としつつ、わが国の死亡状況にあわせて改良を行っ