

研 究 論 文

確率推計による将来人口推計 の不確実性の評価について

石 井 太

I はじめに

国立社会保障・人口問題研究所「日本の将来推計人口（平成14年1月推計）」では、推計手法として「コーホート要因法」が用いられている（国立社会保障・人口問題研究所2002）。これは、人口のコーホート変動要因である、出生率、死亡率（生残率）、人口移動（国際人口移動）、ならびに出生性比に関して仮定を設定し、これらのコーホート変動要因に従って人口を変化させる決定論的な推計である。平成14年推計では、出生率仮定を「高位・中位・低位」の3通り設定し、仮定の変動に起因する人口推計結果の「不確実性」（uncertainty）を表現している。このように複数の仮定を設定する方法は、人口推計結果の不確実性を表現する有力な方法の一つであるが、一方で、この方法では単に複数の推計結果が提示されるだけであり、複数の推計結果がそれぞれ起こりうる確率や、推計結果の信頼区間などが示されていないという指摘もある。本研究では、確率推計による将来人口推計の不確実性の評価について、平成14年1月推計をベースとしたシミュレーションを実行して検討を行った。

II 人口推計モデルと確率推計

1. 確率推計の位置づけ

本研究においては、人口推計における確率推計として、以下の2種類を区別して考えることとする。

- A. 推計に内在する誤差などに起因する推計結果の不確実性を示す確率推計
- B. マイクロシミュレーションなど推計過程に確率的プロセスを持つ確率推計

Aは推計における基礎データ・モデルなどが持つ誤差に起因して推計結果に揺らぎが生じることを表現する確率推計であり、推計結果の不確実性を信頼区間の表示などにより表

現することを目的とするものである。一方、Bは、推計過程に決定論的でない確率性を導入する事により、例えば、平均値でしか表されていない推計結果に対してその分布を示すことなどを目的とするものである。

Aに関する先行研究としては、Lutz et al. (1998)、Lee (1998)、Lee and Carter (1992) など多くのものが挙げられる。また、わが国における研究としては、年金財政を検討する目的から確率推計を将来人口推計に応用した研究として、鈴木 [等] (2003)、北村・中嶋 (2004) などがある他、結婚と出生に関する同時方程式モデルを構築することにより、確率的手法に基づいた出生率将来推計を行った加藤 (2005) などを先行研究として挙げることができる。

一方、Bに関する確率推計の一例がマイクロシミュレーションである。一定の範囲でまとめられた、集団における特性値の平均しか推計できないマクロシミュレーションに対し、個別に履歴を作っていくことにより特性値の分布などの推計を行うことができる点がマイクロシミュレーションを用いる利点である。将来人口推計に関連するわが国での先行研究の例として、世帯推計に応用したInagaki (2005)、府川 (2005)、基礎年金の被保険者期間推計に応用した石井 (2000) などを近年のものとして挙げる事ができる。マイクロシミュレーションを用いることにより、マクロシミュレーションでは示すことが難しかった特性値に関する情報を提供することが可能となる。

また、人口問題研究所での先行研究の例として、昭和55～58年特別研究「出生力の生物人口学的分析」(厚生省人口問題研究所 1984) では、出生力決定に影響を及ぼす人口学的、生物統計学的要因との関連に関するモデルを構築し、マイクロシミュレーションによって出生力と諸要因の関係を把握する研究が行われている。また、平成元～3年度特別研究「家族形成モデルの開発と応用に関する研究」(厚生省人口問題研究所 1990) では、婚姻過程やパリティ別出生率などをモデルに取り入れたマイクロシミュレーションによる出生率予測モデルの構築が試みられており、このような試みは決定論的な将来人口推計の際の仮定設定の検討にも有用なものと考えられる。

このような有効性をもつマイクロシミュレーションであるが、デメリットとしてシミュレーションに必要となる遷移確率などの基礎となるデータの作成が困難なことが多い点が挙げられる。しかしながら、近年、縦断調査などこれらの遷移確率を得ることに適した統計調査なども実施されるようになってきていることから、マイクロシミュレーションのさらなる活用の可能性も広がってきているといえる。

本研究では、以下、Aの確率推計を中心として論じることとする。

2. 不確実性の発生源とその評価方法

人口推計モデルにおいて、不確実性を確率推計として示す研究は数多く存在し、確率推計の位置づけに関しても研究がなされてきている。本稿ではまずLee (1998) を基礎として、人口推計の不確実性へのアプローチについてまとめる。

Lee (1998) は、人口推計における不確実性の発生源として以下の8点を挙げている。

1. プロセス分岐の不確実性
2. 経時的变化
3. データの正確性
4. モデル特性
5. パラメータ推定
6. 将来の政策決定
7. 社会の構造変化
8. カタストロフ

さらに Lee (1998) では人口推計モデルに内在するこれらの不確実性の取扱いについて以下のように述べており、本研究が対象とする平成14年推計の出生率・生命表仮定に関する不確実性においても、これと概ね同様の取扱いをすることが考え方の基礎にある。「1. プロセス分岐の不確実性」「8. カタストロフ」については推計モデルの対象外の問題である。「6. 将来の政策決定」「7. 社会の構造変化」については、推計期間内における保健・福祉水準の変化などは実績の死亡率データ等が内包しており、それに対応する分は既に推計モデルに反映されているといえる。ただし、今後の変化度合が異なる可能性もあり、その不確実性はなお存在する。「4. モデル特性」については、モデルの説明力とモデル選択の問題がある。モデルの説明力については、過去のデータと完全に適合をしているわけではなく、将来についてもその不確実性は存在する。また、モデル選択に関しては、選択した時点でそれ以外の可能性は表現されないため、モデル選択の誤りに基づく不確実性は存在する。「5. パラメータ推定」についても、推定を行った時点でそれ以外の可能性は表現しえないため、不確実性は存在する。「2. 経時的变化」は6・7と関係する部分もあるが、今後の経時的变化の不確実性は過去の変動以上のものは存在するといえる。「3. データの正確性」については、例えば生命表モデルに関してはデータソースが完全生命表及び簡易生命表であることから大きな影響はないと考えてよい一方、出生モデルの推計に出生動向基本調査などの標本調査を用いるような場合にはその影響に関する考慮が必要な問題となろう。

また、予測誤差の評価にあたっては、誤差間の相関も考慮が必要な問題である。Lee (1998) は、年齢間の予測誤差は完全な相関、出生率・死亡率・人口移動の予測誤差は無相関と考えてよいが、予測誤差の時系列間の相関には慎重な検討が必要としている。

次に、不確実性の表現方法については、平成14年推計でも用いられた「高位－中位－低位」のようなシナリオベースの表現方法、種々の専門家の意見を並列することによる表現方法などがあるが、それらは必ずしも不確実性を確率的には表現していない。推計結果の信頼区間などの確率的表現を行うためには確率推計が必要となる。この確率推計の手法については、過去の推計結果を分析する Ex Post Analysis で得られるものを利用する方法や、Lutz et al. (1998) などによる専門家調査結果から仮定数値の分布を作ってそれを利用するもの、生命表推計の Lee-Carter 法 (Lee and Carter 1992) で行われているような時系列分析を応用する方法など、幾つもの方法が研究されている。

本研究で検討する確率推計については、先ほど挙げられた種々の不確実性要因に対する取扱いを基礎としつつも、それぞれの要因を個々に取り扱うのではなく、専門家の予測の不確実性がこれらの不確実性を全体として表現していると考えて各種前提の確率分布を設定し、平成14年1月推計をベースとしたシミュレーションを実行するものである。

3. シミュレーションの全体像

本研究では、コーホート要因法による将来人口推計の前提のうち、出生率仮定・生残率仮定について確率的に仮定設定を行い、これに基づくシミュレーションを実行して将来人口推計結果の信頼区間などを作成する確率推計を行う。各前提に用いる分布やそれに従う仮定値の設定方法などについては後に詳細に述べるが、ここではまず、シミュレーションの全体像について述べる。

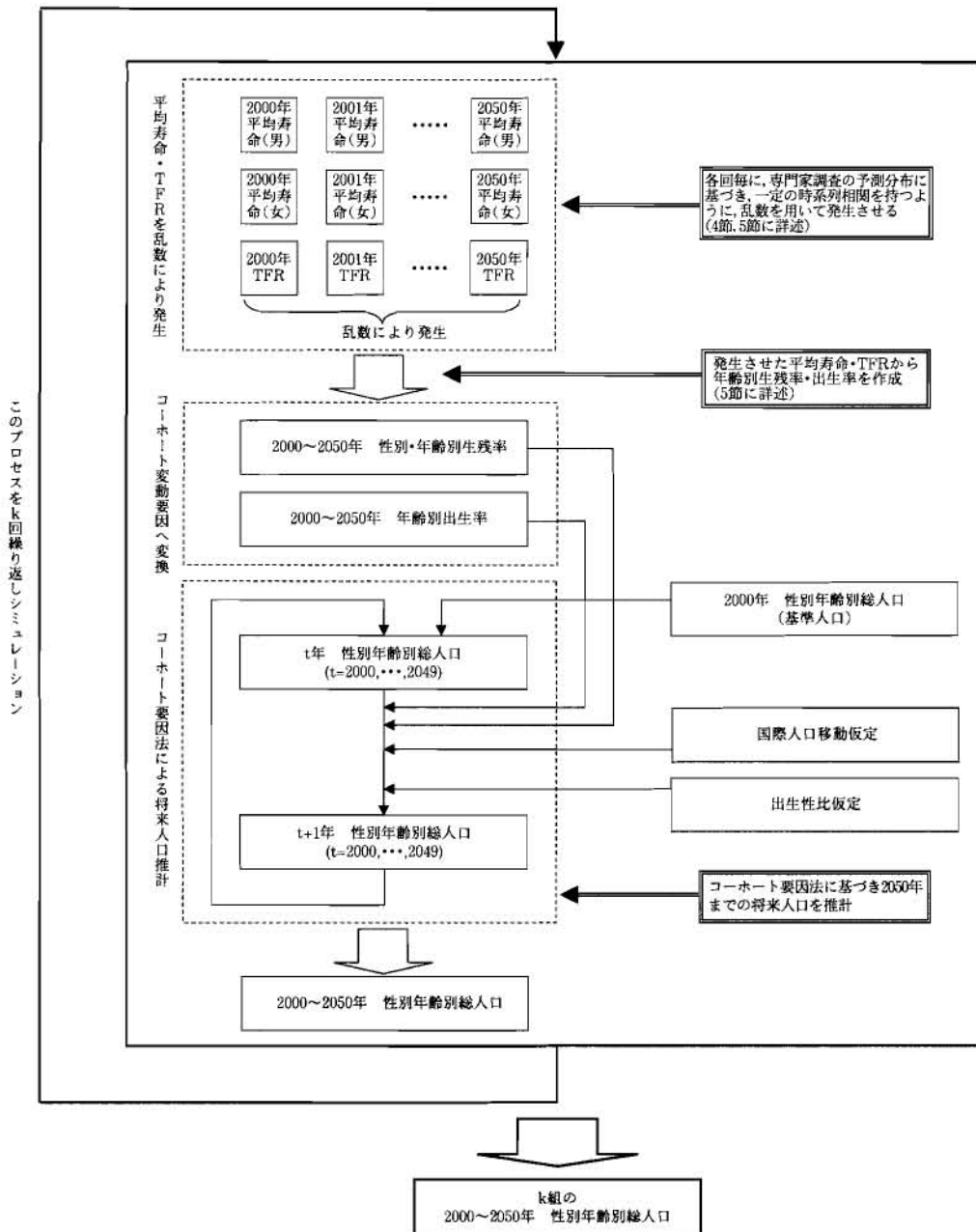
図1は本研究で行ったシミュレーションの全体像を表したものである。大きい四角で囲われているのが、繰り返しにおける1回のプロセスを表し、これを複数回(k 回)実行して k 組の将来人口推計結果を作成する。本研究ではこのプロセスを1000回($k=1000$)実行した。

次にプロセス内部について述べる。まず、一番上の点線で囲まれた部分に示されているように、推計期間である2000～2050年の男女別平均寿命とTFRを乱数を用いて発生させる。このとき、発生する平均寿命及びTFRは後述する専門家調査の予測分布に従うものとし、かつ、プロセス内では一定の時系列相関を持つようにする。これらの具体的方法論については4節、5節でより詳しく述べる。

次に、発生させた平均寿命・TFRを、5節で述べる方法を用いて、年次別、年齢別生残率・出生率へと変換する。国際人口移動仮定・出生性比は平成14年推計の前提を固定して用いることとすると、これによりコーホート要因法による将来人口推計を行うための前提が1組整うこととなる。そこで、これらの前提に基づき、コーホート要因法により将来人口推計を実行して2000～2050年の性別・年齢別総人口を得る。したがって、これらを k 回繰り返すことにより、 k 組の将来人口推計結果が得られることとなるわけである。

シミュレーションの全体像は以上に述べた通りであるが、以下では、シミュレーションに用いられた平均寿命やTFRの予測分布や仮定値の設定方法などについてより詳細に述べることとする。

図1 シミュレーションの全体像



4. 専門家調査の予測値の分布

本研究では、専門家の予測の不確実性として、「少子化に関する専門家調査」に基づく専門家全体の予測値の分布を用いることとした（守泉 2004）。同調査では、2050年における男女別平均寿命の予測値、2025年における合計特殊出生率の予測値が調査されている。平均寿命の予測値の分布を示したものが、図2、図3の実線のグラフである。

図2 平均寿命の予測値の分布(男)

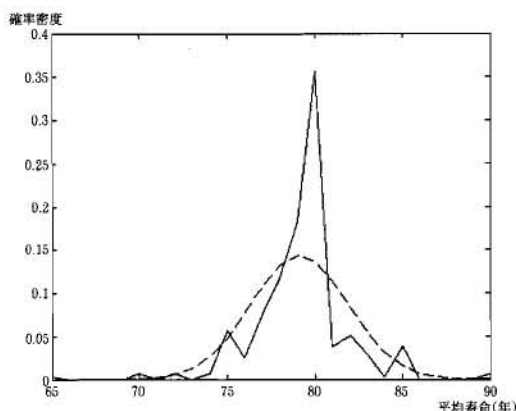
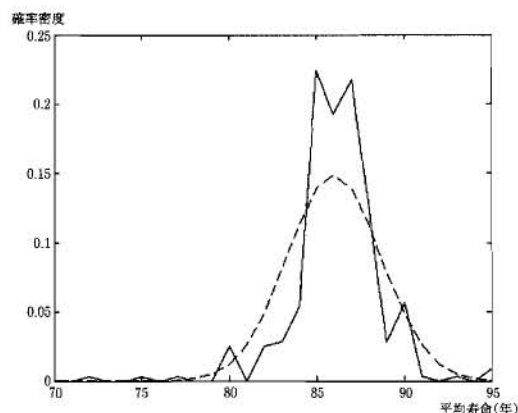


図3 平均寿命の予測値の分布(女)



比較のため、同じ図に専門家の予測と同一の平均、標準偏差を持つ正規分布を点線で示した。これによると、専門家調査の予測値は平均値を概ね中心として左右に分布しているものの、同一標準偏差の正規分布と比較した場合、平均値の周辺により集中した分布となっているのがわかる。従って、確率推計に用いる平均寿命の分布については、同一平均・同一変動係数の正規分布を用いるよりも専門家調査の予測値の分布を用いる方が好ましいと考えられるが、一方で専門家調査の予測値はサンプルサイズがそれほど大きくないこともあり、そのまま用いると推計結果が不安定になることから、平滑化などが必要と考えられる。そこで、本研究においては、正規分布をカーネル関数としたカーネル密度推定を行って得た分布を推計に用いることとした。カーネル密度推定とは、得られたサンプルの値 $\{x_1, x_2, \dots, x_n\}$ から、確率密度関数 $f(x)$ を、

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$$

により推定するものである（ジェフリー 1999）。ただし、 $K(x)$ ：カーネル関数は、ここでは標準正規分布の確率密度関数とする。ここで、 h はバンド幅と呼ばれる。 h については、推定の対象となる分布が正規分布に近い場合、 σ を標準偏差として、

$$h_0 = \left(\frac{4}{3n}\right)^{\frac{1}{5}} \sigma$$

とするものがよく用いられる。しかしながら、今回の平均寿命に関する専門家調査の予測値の分布は正規分布とは形状が大きく異なっており、このバンド幅を用いた推定では平滑化後の分布のスムーズさにやや問題が見られたことから、 $h = 2h_0$ を用いることとした。平滑化後の密度関数推定結果を示したものが、図4、図5である。

図4 平均寿命の予測値の分布(平滑化後・男)

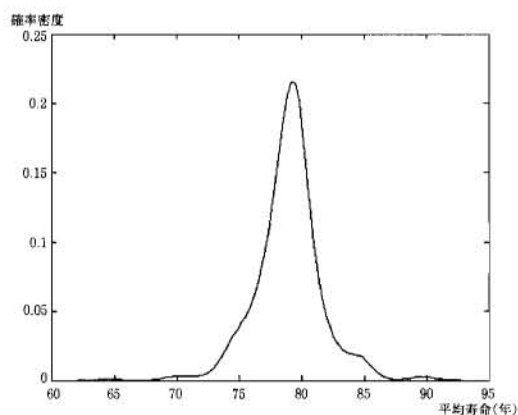
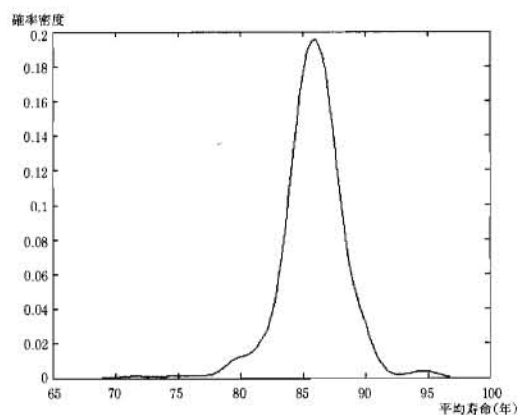


図5 平均寿命の予測値の分布(平滑化後・女)



次に、2025年の TFR の予測値の分布を示したものが、図6の実線のグラフである。平均寿命のものと同様に、同一の平均、標準偏差を持つ正規分布を点線で示している。TFR についても平均寿命の分布と同様、正規分布をカーネル関数としたカーネル密度推定を行ったが、TFR の予測値の分布は比較的正規分布に近いことから、バンド幅 h については、 $h = h_0$ とした。平滑化後の密度関数推定結果を示したものが、図7である。

図6 TFR の予測値の分布

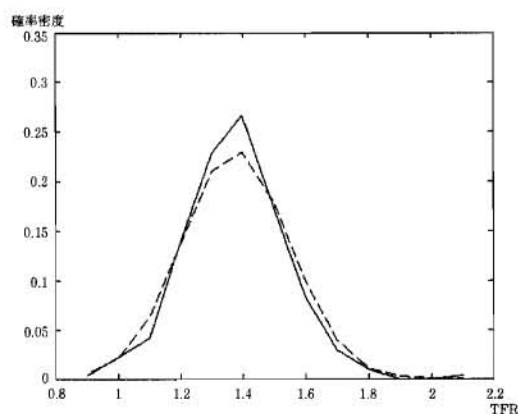
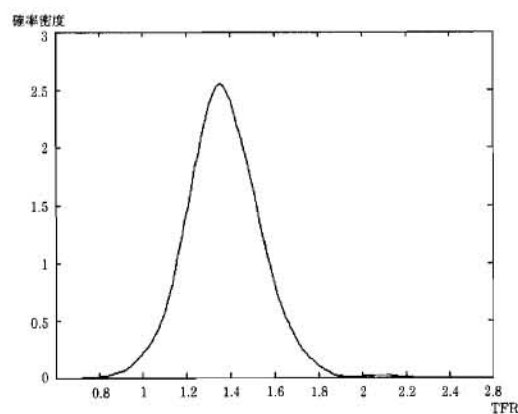


図7 TFR の予測値の分布(平滑化後)



これらについて、平滑前後における分布の特性値の変動を見たものが表1である。これによれば、どの分布についても平滑化によりやや変動係数が増大しているが、本研究では平滑後の分布を専門家の予測値の分布の仮定値として利用することとした。

表1 予測値の分布の特性値

平均寿命（男）	平滑化前	平均値	79.29
		標準偏差	2.59
		変動係数	3.27%
	平滑化後	平均値	79.13
		標準偏差	2.77
		変動係数	3.50%
平均寿命（女）	平滑化前	平均値	86.15
		標準偏差	2.48
		変動係数	2.87%
	平滑化後	平均値	85.99
		標準偏差	2.66
		変動係数	3.10%
TFR	平滑化前	平均値	1.38
		標準偏差	0.16
		変動係数	11.46%
	平滑化後	平均値	1.38
		標準偏差	0.17
		変動係数	12.30%

なお、以下では、これら専門家調査の予測値の分布について、その分布関数・密度関数・平均値・変動係数を、

$$\bar{F}_{2050}^{e_{0m}}(x), \bar{f}_{2050}^{e_{0m}}(x), \bar{\mu}_{2050}^{e_{0m}}, \bar{C}V_{2050}^{e_{0m}} \quad (2050\text{年平均寿命・男})$$

$$\bar{F}_{2050}^{e_{0f}}(x), \bar{f}_{2050}^{e_{0f}}(x), \bar{\mu}_{2050}^{e_{0f}}, \bar{C}V_{2050}^{e_{0f}} \quad (2050\text{年平均寿命・女})$$

$$\bar{F}_{2025}^{TFR}(x), \bar{f}_{2025}^{TFR}(x), \bar{\mu}_{2025}^{TFR}, \bar{C}V_{2025}^{TFR} \quad (2025\text{年 } TFR)$$

で表すこととする。

5. 確率推計における各種前提の設定

次に、確率推計における各種前提の設定方法について述べる。

我々は、前節で述べた専門家調査の予測値分布を利用して、2050年までの各年における生命表及び年齢別出生率に関する確率的な仮定をおく必要がある。そこで、以下のような基本的考え方の下にこれを実行する。

1. 各年における平均寿命・TFRの分布は、専門家調査による予測分布に平均・変動係数の線形変換のみを施したものとする

2. 各年における分布の平均値は、平均寿命は平成14年推計、TFR は平成14年推計の中位推計に一致
3. 各年における分布の変動係数は、年次間で線形補間を行うことにより設定
4. 各年における年齢別死亡率・出生率は、平均寿命・TFR にあわせて設定
5. 年次間における分布の相関は、平均寿命・TFR の過去の実績値の自己相関係数を用いて設定

なお、国際人口移動に関しては専門家調査の結果がないため、本研究では確率的な前提設定を行わず、平成14年推計の仮定値を固定して用いることとした。

さて、平均寿命については、先述の専門家調査の予測分布が2050年における分布となるため、2000年の変動係数を0として2050年の変動係数との間で線形補間を行い、各年の平均寿命の変動係数を設定する。具体的には、 $CV_n^{\dot{e}_0^*} (* = m, f)$ を n 年の平均寿命の変動係数とすると、

$$CV_n^{\dot{e}_0^*} = (n-2000)/50 \cdot CV_{2050}^{\dot{e}_0^*}$$

となる。これにより、 $\mu_n^{\dot{e}_0^*}$ を平成14年推計の n 年の平均寿命として、 n 年の平均寿命の分布関数は、

$$F_n^{\dot{e}_0^*}(x) = \bar{F}_n^{\dot{e}_0^*} \left(\frac{x - \bar{\mu}_{2050}^{\dot{e}_0^*}}{\bar{\mu}_{2050}^{\dot{e}_0^*} \cdot CV_{2050}^{\dot{e}_0^*}} \cdot \mu_n^{\dot{e}_0^*} \cdot CV_n^{\dot{e}_0^*} + \mu_n^{\dot{e}_0^*} \right)$$

となる。

これらの平均寿命の変動に対応し、年齢別死亡率などの生命関数を得るため、平成14年推計における2050年将来生命表の中央死亡率 m_x を基礎とし、 $-0.1 \leq c \leq 0.1$ (0.01刻み) である c に対し、

$$\dot{e}_0^* = (1+c)\dot{e}_0$$

$$\log m_x^* = k \log m_x$$

となる k を求めた。この c と k の関係が2050年以外の年にも成立すると仮定すれば、平均寿命の変動に対応した各年の年齢別中央死亡率を得ることができ、従って生命関数を作成することが可能となる。

次に問題となるのは、年次間における分布の相関である。前章において、Lee (1998) では年次間の相関について慎重な検討が必要であることが指摘されていたことを述べた。そこでこの取扱いについて次に述べる。

一般に年次間の相関については、

1. 年次間の分布が完全に独立であるケース
2. 年次間の分布が完全に相関しているケース

の2ケースを対極的なものとして考えることができる。このうち、1のケースでは各年の仮定値の分布の標準偏差が仮に σ であり、 n 年間の推計を行ったとした場合、平均的な仮定値の分布の標準偏差は $\sigma/\sqrt{n}$ に減少してしまうこととなり、確率的な仮定値の変動を小さく見る危険性がある。一方、2のケースでは平均的な仮定値の分布の標準偏差は σ となり、変動を過小に評価することはない。このため、専門家調査の分布を用いる場合には2のケースのような仮定設定により確率推計を実行することが多い。

しかしながら、この方法では仮定値の時系列での動きが大きい制限を受けることとなり、多様な時系列変動を仮定値として含むことが許容されない。マルコフ連鎖などにより仮定値を逐次的に発生させるとこの問題は克服できるが、各年次における仮定値の分布に専門家調査などの情報を用いることが難しい (Lee 1998)。

そこで、本研究では専門家調査による予測値の分布情報を活用しつつ、各年次間の時系列に一定の相関を持たせる観点から、以下のような手法を用いて確率推計を実施した。

一般に、 $X_1, X_2, \dots, X_n$: 確率変数、 $F(x_1, x_2, \dots, x_n)$: 連続同時分布関数、 $F_1, F_2, \dots, F_n$: 連続周辺分布の間に、以下の Sklar's theorem が成立することが知られている (Nelsen 2006)。

定理 (Sklar) $F_1, F_2, \dots, F_n$: 連続周辺分布をもつ n 変量連続同時分布関数を F としたとき、関数 C が unique に存在して以下を満たす。

$$Pr(X_1 \leq x_1, \dots, X_n \leq x_n) = F(x_1, x_2, \dots, x_n) = C(F_1(x_1), (F_2(x_2), \dots, F_n(x_n)))$$

この関数 C はコピュラと呼ばれ、周辺分布と同時分布の関係を示すものであり、金融実務等で利用されている (戸坂・吉羽 2005)。これにはいくつかの典型的なものが知られているが、本研究では以下に述べる正規コピュラを用いた。

正規コピュラとは、 $X_1, X_2, \dots, X_n$: 確率変数とし、相関行列 Σ を持つ n 変量標準正規分布の分布関数を $\Phi_n(x_1, x_2, \dots, x_n; \Sigma)$ としたとき、周辺分布は1変量正規分布であることから、1変量標準正規分布の分布関数を $\Phi_1(x)$ と書けば、Sklar's theorem から C : コピュラが存在して、

$$\begin{aligned} Pr(X_1 \leq x_1, \dots, X_n \leq x_n) &= \Phi_n(x_1, x_2, \dots, x_n; \Sigma) \\ &= C(\Phi_1(x_1), \Phi_2(x_2), \dots, \Phi_n(x_n)) \end{aligned}$$

を満たす。これを正規コピュラと呼ぶ。

本研究では、この正規コピュラにより時系列間の分布に相関を持たせることとした。具体的には以下の手順によった。

1. 相関行列 Σ の51変量正規分布に従う乱数

$$(z_1, \dots, z_{51})$$

を発生

2. 各変数を標準正規分布の分布関数 $\Phi_1(x)$ を用いて変換し,

$$(\Phi_1(z_1), \dots, \Phi_1(z_{51}))$$

を得る

3. これに各年の平均寿命分布関数の逆関数を作用させ,

$$(F_{2000}^{\hat{e}_0^*}{}^{-1}(\Phi_1(z_1)), \dots, F_{2050}^{\hat{e}_0^*}{}^{-1}(\Phi_1(z_{51})))$$

として各年の平均寿命を得る

ここで、相関行列 Σ は、

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{50} \\ \rho & 1 & \rho & \dots & \rho^{49} \\ \rho^2 & \rho & 1 & \dots & \rho^{48} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{50} & \rho^{49} & \rho^{48} & \dots & 1 \end{bmatrix}$$

とし、 ρ は、1948～2004年の男女別平均寿命の前年との自己相関係数とした ($\rho = 0.99681$ (男), 0.99662 (女))。この相関行列は、周辺分布毎に逆変換を施して得た多変量正規分布が AR(1) モデルに一致することを意味する。

なお、ここで一点注意が必要であるのは、この自己相関係数は正確にはこの多変量正規分布における二変数間の相関係数であって、平均寿命の時系列において1年差を取った自己相関係数とは異なることである。しかしながら、予測値の周辺分布とは異なり、変数間の相関については線形関係に概ね近い関係がある限り大きく異なるものとはならない。実際に、上に述べた相関係数を持つ2変量正規分布から、2変量の平均寿命への変換を行うシミュレーションを10,000回実行し、相関係数を計算したところ、変換後の相関係数は、 $\rho' = 0.99624$ (男), 0.99629 (女) となり、変換前と概ね同じ値となっていることが確認できた。従って、本研究ではこのような方法により確率推計を行うことで適切な結果が得られると考えられるが、予測値の分布が多峰性を持つ場合など、正規分布から予測値の分布への変換が相関係数に影響を及ぼす恐れがあると考えられる場合には、別途検討が必要となる場合もあるので注意が必要である。

出生率についても概ね同様の手順によるが、TFRについては専門家調査の予測値の分布が2025年であることから、 CV_n^{TFR} を n 年の TFR の変動係数とし、

$$CV_n^{TFR} = (k - 2000) / 25 \cdot \bar{CV}_{2025}^{TFR}$$

として設定を行った。年齢別出生率は、各年次において出生率の年齢パターンを固定し、

TFR の変動率をそのまま年齢別出生率の変動率とした。また、相関係数行列に用いる ρ は、1947～2004年の TFR の前年との自己相関係数とした ($\rho = 0.97725$)。なお、平均寿命で確認したのと同様に、変換後の相関係数のシミュレーション (10,000回) を行うと $\rho' = 0.97671$ と、やはり概ね変換前と同じ値であることが確認できた。

Ⅲ 推計結果

1. 前提

前章で述べたシミュレーションの全体像及び各種仮定設定法に従って、将来人口推計を繰り返し (1000回) 実施するシミュレーションを行った。本章ではそれらの推計結果について述べることにする。

表2は、各仮定値の平成14年推計における値と行ったシミュレーションの分布の特性値、設定に使用した専門家調査による分布 (平滑化後) を比較したものである。この表から、シミュレーション結果の分布の平均値は概ね平成14年推計のものに一致しているとともに、変動係数も設定に用いた専門家調査と概ね一致しており、確率的な仮定設定が適切に行われていることが確認できる。

表2 各種前提とシミュレーション結果の比較

年 次			2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
平均寿命 (男)	平成14年推計		77.64	78.11	78.62	79.05	79.43	79.76	80.06	80.32	80.55	80.76	80.95
	Simulation	平均値	77.64	78.10	78.61	79.05	79.43	79.77	80.07	80.34	80.58	80.80	80.97
		標準偏差	0.00	0.29	0.57	0.85	1.14	1.39	1.72	2.02	2.35	2.73	3.05
		変動係数	0.00%	0.37%	0.72%	1.07%	1.44%	1.75%	2.15%	2.51%	2.92%	3.37%	3.77%
	専門家調査 (平滑化後)	平均値											79.13
		標準偏差											2.77
		変動係数											3.50%
平均寿命 (女)	平成14年推計		84.62	85.20	85.90	86.51	87.05	87.52	87.93	88.31	88.64	88.94	89.22
	Simulation	平均値	84.62	85.21	85.92	86.54	87.09	87.56	87.99	88.38	88.73	89.02	89.31
		標準偏差	0.00	0.28	0.56	0.83	1.09	1.35	1.62	1.86	2.16	2.40	2.68
		変動係数	0.00%	0.33%	0.65%	0.95%	1.25%	1.54%	1.84%	2.11%	2.43%	2.69%	3.00%
	専門家調査 (平滑化後)	平均値											85.99
		標準偏差											2.66
		変動係数											3.10%
TFR	平成14年推計	中位	1.36	1.31	1.32	1.35	1.38	1.38	1.38	1.39	1.39	1.39	1.39
		高位	1.36	1.39	1.48	1.56	1.61	1.62	1.62	1.63	1.63	1.63	1.63
		低位	1.36	1.22	1.14	1.12	1.11	1.11	1.10	1.10	1.10	1.10	1.10
	Simulation	平均値	1.36	1.31	1.32	1.35	1.37	1.38	1.39	1.39	1.39	1.39	1.39
		標準偏差	0.00	0.03	0.07	0.10	0.14	0.17	0.21	0.25	0.27	0.30	0.34
		変動係数	0.00%	2.54%	4.97%	7.34%	10.04%	12.47%	14.82%	17.71%	19.50%	21.90%	24.48%
	専門家調査 (平滑化後)	平均値						1.38					
		標準偏差						0.17					
		変動係数						12.30%					

次に、各仮定値ごとに、シミュレーション結果の分布を示す。

図8、図9は、各年における平均寿命 e_0 の90%、95%信頼区間及び分布の平均値・中央値を表したものである。

図8 各年における平均寿命 e_0 の信頼区間等(男)

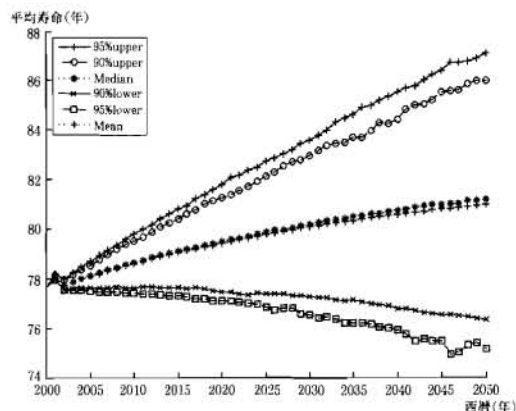
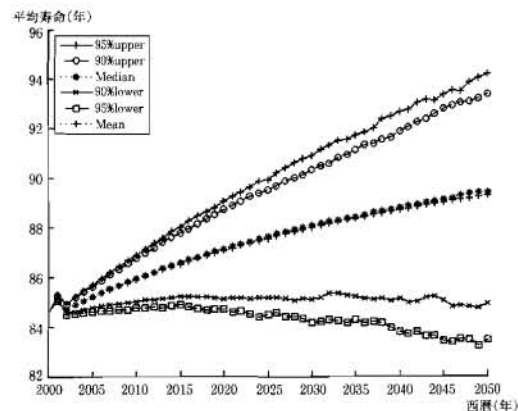


図9 各年における平均寿命 e_0 の信頼区間等(女)



これらの図によれば、2050年における平均寿命の信頼区間は、男では95%信頼区間は [75.1, 87.1] と11.9年の幅となっている。一方、女では、95%信頼区間は [83.5, 94.2] と10.7年の幅となっている。

これらについては、各年毎には専門家調査の分布を線形変換したものを用いているため、正規分布とは異なる分布となっている。5年毎の分布を示したものが図10、図11、その拡大図について示したものが図12、図13である。特に男子の分布においては専門家調査の分布が非対称となっていることがシミュレーション結果に反映されていることが確認できる。

図10 各年における平均寿命 e_0 の分布(男)

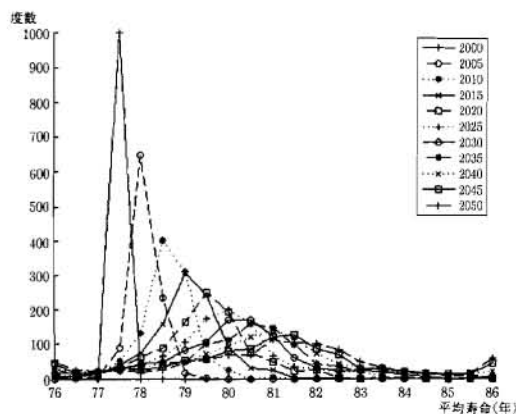


図11 各年における平均寿命 e_0 の分布(女)

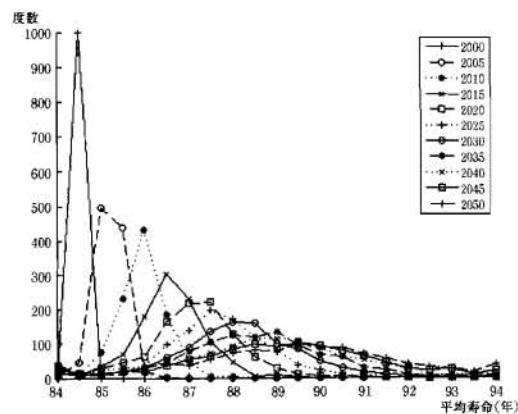


図12 各年における平均寿命 e_0 の分布(男・拡大図)

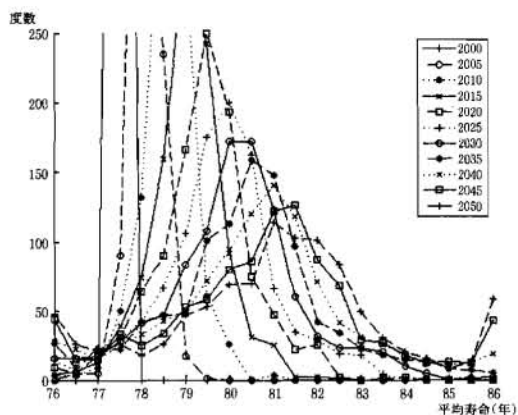


図14 平均寿命 e_0 の軌跡(男)

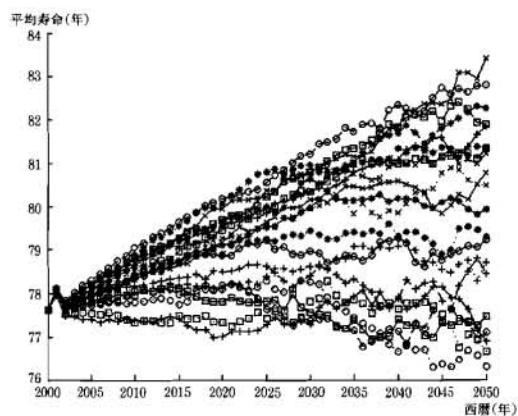


図13 各年における平均寿命 e_0 の分布(女・拡大図)

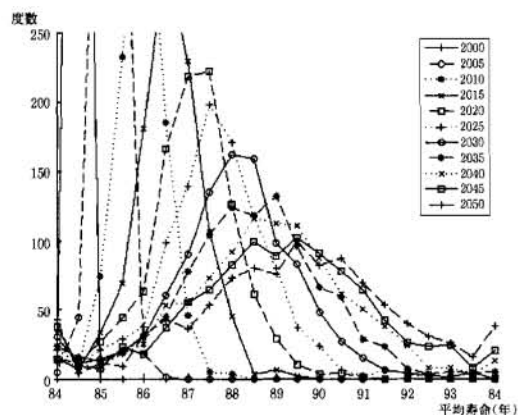
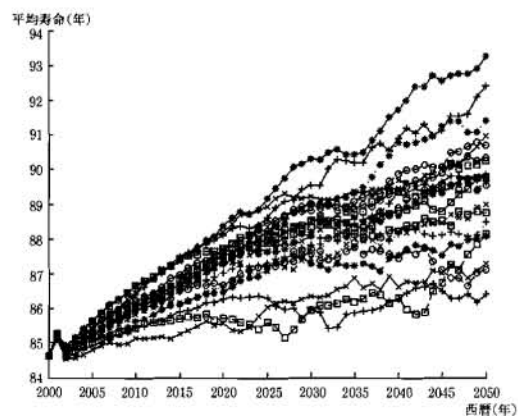


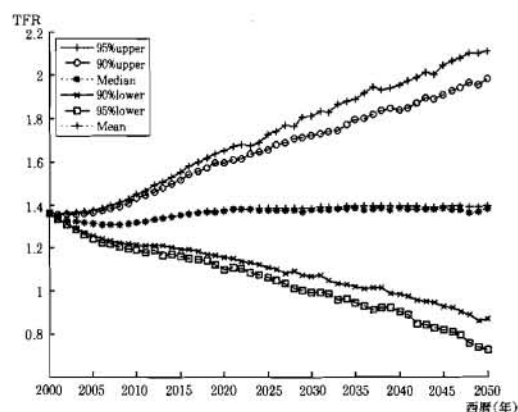
図15 平均寿命 e_0 の軌跡(女)



時系列間の相関状況を確認するため、いくつかのシミュレーション結果の軌跡を例示したものが、図14、図15である。これによれば、平均寿命に関しては自己相関係数が高いことから、一旦高い値をとると、その後も概ね高い軌跡を描くなどの傾向を持つものとなっているが、一方で、完全な相関を持ったものに比べ、より自由度が高く多様な軌跡を含みうるシミュレーション結果となっていることが分かる。

次に、TFRについても同様のものを見てみることにする。図16は、各年におけるTFRの90%、95%信頼区間及び分布の平均値・中央値を表したものである。

図16 各年におけるTFRの信頼区間等



2050年における合計特殊出生率の信頼区間は、95%信頼区間が $[0.72, 2.11]$ と1.39の幅となっている。平成14年推計における2050年における低位推計、高位推計のTFRがそれぞれ1.10、1.63であることから、95%信頼区間は両者の幅より大きいものとなっていることが分かる。5年毎の分布を示したものが図17、その拡大図について示したものが図18である。

図17 各年におけるTFRの分布

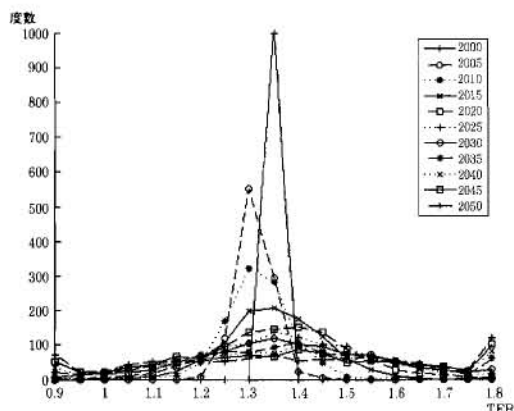
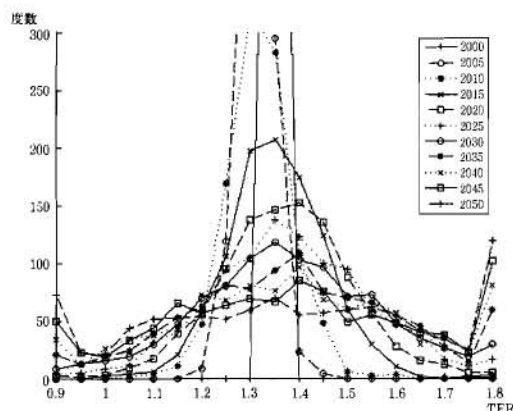
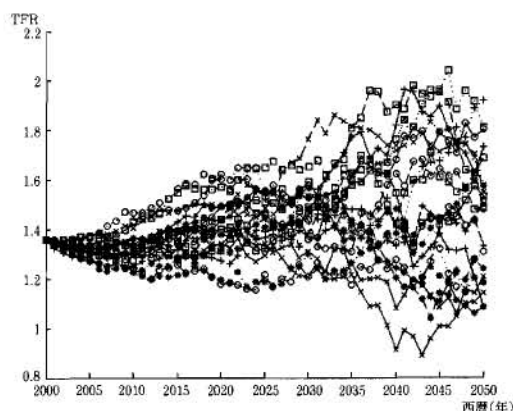


図18 各年におけるTFRの分布(拡大図)



また、時系列間の相関状況を確認するため、いくつかのシミュレーション結果の軌跡を例示したものが、図19である。TFRに関しては平均寿命に比べ自己相関係数が低いため、より自由度が高い軌跡を含むシミュレーション結果となっている。

図19 TFRの軌跡



2. 総人口・年齢別人口割合の推計結果

次にこれらの前提に基づいて人口推計を行った結果について示す。まず、総人口と年齢別人口割合について、5年毎の95%信頼区間及び分布の標準偏差・変動係数を中位推計結果と比較したものが表3である。総人口については、90%、平均値、中央値と併せたものを図20に示している。

これによれば、2050年の95%信頼区間は $[8984\text{万人}, 11224\text{万人}]$ となっている。このシミュレーションにおいては生命表の前提も変動させていることから単純な比較はできないものの、平成14年推計の2050年における総人口は、低位推計で9203万人、高位推計で10825万人となっており、この95%信頼区間は両者の幅より大きいものとなっている。また、変動係数を見ると、2025年では1.4%であるのに対して、2050年は5.3%となっており、

表3 総人口・年齢別人口割合の推計結果

年次		2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
総人口(万人)	中位推計結果	12,693	12,771	12,747	12,627	12,411	12,114	11,758	11,360	10,934	10,496	10,059
	95%上限	12,693	12,788	12,810	12,755	12,627	12,436	12,209	11,982	11,731	11,462	11,224
	95%下限	12,693	12,752	12,682	12,492	12,184	11,795	11,322	10,791	10,220	9,614	8,984
	標準偏差	0	9	32	68	113	165	226	294	369	450	536
	変動係数	0.000	0.001	0.003	0.005	0.009	0.014	0.019	0.026	0.034	0.043	0.053
年少人口割合 (14歳以下割合)	中位推計結果	0.146	0.139	0.134	0.128	0.122	0.116	0.113	0.111	0.110	0.109	0.108
	95%上限	0.146	0.140	0.138	0.136	0.134	0.132	0.133	0.135	0.142	0.145	0.149
	95%下限	0.146	0.138	0.130	0.121	0.110	0.101	0.092	0.086	0.081	0.077	0.068
	標準偏差	0.000	0.001	0.002	0.004	0.006	0.008	0.010	0.012	0.015	0.017	0.020
	変動係数	0.000	0.004	0.014	0.031	0.051	0.071	0.091	0.113	0.136	0.159	0.184
生産年齢人口割合 (15～64歳割合)	中位推計結果	0.681	0.662	0.641	0.612	0.600	0.597	0.592	0.580	0.558	0.544	0.536
	95%上限	0.681	0.663	0.644	0.618	0.609	0.608	0.606	0.595	0.575	0.564	0.559
	95%下限	0.681	0.662	0.638	0.606	0.591	0.586	0.578	0.563	0.539	0.523	0.511
	標準偏差	0.000	0.000	0.001	0.003	0.005	0.006	0.007	0.008	0.009	0.010	0.012
	変動係数	0.000	0.001	0.002	0.005	0.008	0.010	0.012	0.014	0.016	0.019	0.022
老年人口割合 (65歳以上割合)	中位推計結果	0.174	0.199	0.225	0.260	0.278	0.287	0.296	0.309	0.332	0.347	0.357
	95%上限	0.174	0.199	0.227	0.263	0.285	0.297	0.310	0.330	0.359	0.381	0.399
	95%下限	0.174	0.198	0.224	0.256	0.272	0.277	0.282	0.291	0.307	0.315	0.316
	標準偏差	0.000	0.000	0.001	0.002	0.003	0.005	0.007	0.010	0.013	0.017	0.021
	変動係数	0.000	0.001	0.004	0.007	0.012	0.018	0.025	0.032	0.039	0.048	0.058

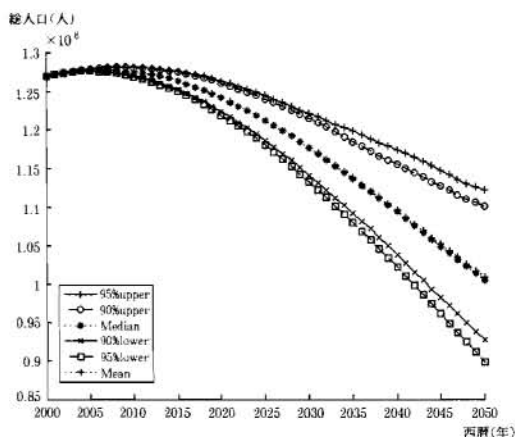
推計期間後半での不確実性の増大が大きくなることが分かる。

一方、年齢別人口割合はそれぞれで異なった動きが観察される。年少人口割合については、出生率仮定の変動の影響を受け、2050年の95%信頼区間は[6.8%, 14.9%]、変動係数は18.4%とかなり大きいものとなっており、年少人口割合の推計結果自体を重視する場合には一定の留意が必要といえよう。しかしながら、中位推計における年少人口割合は2050年に10.8%まで減少することから、標準偏差では2%ポイントに留まっているともいえる。

生産年齢人口割合については、これらの3区分の中では最も安定した結果といえる。2050年での95%信頼区間は[51.1%, 55.9%]、標準偏差は1.2%ポイント、変動係数では2.2%であり、生命表仮定や出生率仮定の変動による不確実性が他の年齢別人口割合に比較すると大きくないとの結果が得られた。

一方、高齢化率を示す老年人口割合については、90%、平均値、中央値と併せたものも図21に示したが、年少人口割合ほどではないものの生産年齢人口割合に比べると不確実性は大きいといえる。2050年における95%信頼区間は[31.6%, 39.9%]であり、標準偏差は2.1%ポイント、変動係数では5.8%である。また、総人口同様、推計期間後半での不確

図20 総人口の推計結果の信頼区間



実性の増大が大きいといえる。

これらの結果から分かるように、将来人口推計結果の不確実性は、対象とする人口変数や、年齢階級、年次によって異なった様相を見せており、確率推計による不確実性の評価を参考にすることにより、推計結果の活用に対する見方が広がるといえよう。

3. 生命表・出生率仮定のそれぞれを変化させた場合の推計結果

前節では生命表・出生率仮定を同時に確率的に設定することによりシミュレーションを行ったが、それぞれの仮定の一つだけを確率的に設定した場合、将来人口推計結果がどのような影響を受けるかについても評価を行った。総人口と高齢化率について、その95%信頼区間と中位推計の結果を図に示したものが図22、図23である。

図22 総人口の推計結果の信頼区間

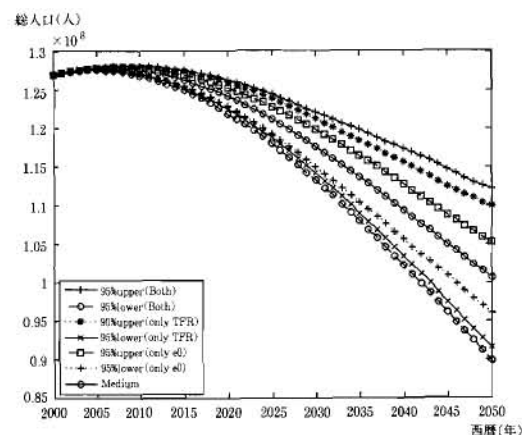


図21 高齢化率の推計結果の信頼区間

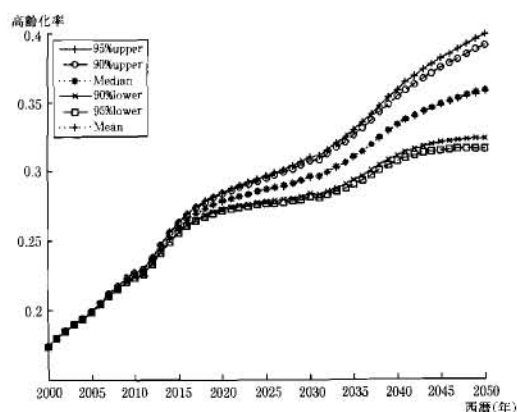
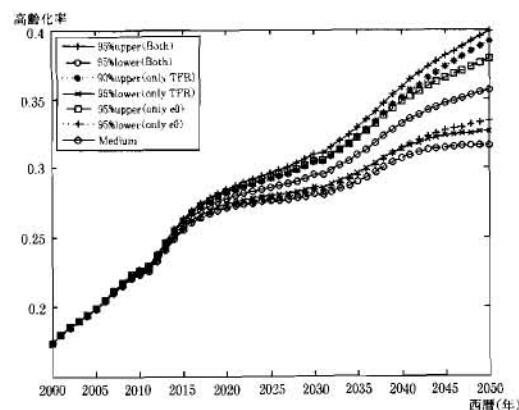


図23 高齢化率の推計結果の信頼区間



総人口について、出生率仮定のみ変動させた場合の95%信頼区間は「9158万人、10987万人」であり、これは平成14年推計における低位推計、高位推計の幅よりやや大きいものとの結果となった。また、生命表仮定のみ変動させた場合の95%信頼区間は「9595万人、10518万人」であり、出生率仮定のみ変動させたものよりやや小さい幅を示した。

また、高齢化率については、出生率仮定のみ変動させた場合の95%信頼区間は「32.6%、39.2%」、生命表仮定のみ変動させた場合の95%信頼区間は「33.4%、37.9%」となった。

4. 総人口がピークとなる年次

平成14年推計において、今後、わが国の人口が継続的な人口減少時代を迎えることが示されたことから、わが国の総人口が実際にいつから減少に転じるかが一般的にも注目されてきた。しかしながら、総人口が増加から減少に転じることは人口学的には一定の過程の中で捉えられるべきものであり、「総人口のピーク年次」を将来人口推計から特定することには人口学的な意味がそれほど大きいとは考えられない。また、推定したとしてもその推計結果の不確実性が高いため、かなり困難な問題でもある。これは、ピーク年次の推定が、総人口の時間微分による導関数の零点を求める操作に相当することに起因するものであるが、その不確実性の高さを一般的にわかりやすく示す試みはこれまで必ずしも多く行われてきたとはいえない面がある。

そこで、前提となる生命表や出生率を確率的に変動させた場合に、将来人口推計における総人口のピーク年次がどのような分布となるかを推計し、その不確実性について評価を行った。平成14年推計の中位推計で総人口がピークとなる年次は2006年であるが、今回の確率推計を用いたシミュレーションによる総人口のピーク年次の分布は表4のとおりである。これによれば総人口が2006年にピークとなる確率は41.1%と約4割でしかなく、ピーク年次を将来推計人口を用いて推定することの困難さが現れているといえよう。なお、今回のシミュレーションでは国際人口移動については確率的な仮定設定を行っていないため、国際人口移動の動向が人口のピーク年次に与える影響をも考慮した場合、ピーク年次の不確実性はさらに高いものとなる可能性があることに注意が必要である。

表4 総人口がピークとなる年次

総人口の ピーク年次	構成割合
2004	1.2%
2005	14.8%
2006	41.1%
2007	26.5%
2008	11.1%
2009	3.8%
2010-	1.5%
	100.0%

IV おわりに

本研究では、専門家調査による予測値の分布を用い、平成14年推計に確率推計を適用することにより、将来人口推計結果の不確実性に関する評価を試みた。本研究で提案した手法は、決定論的に行われた既存の推計結果を基本としつつ、確率的な前提として各年次に対して任意の分布を設定することが可能であり、さらに、一定の時系列相関も表現することができるものである。このため、公式推計が行われた後で、その結果を中心とし、独立に専門家調査の結果を利用して比較的簡明、かつ現実的な時系列軌跡を持つ前提設定を行うことができるという利点がある。

また、本研究においては、実行したシミュレーション結果を用い、総人口、年齢別構成割合等の信頼区間の推定を行った。これらの結果によれば、将来人口推計結果の不確実性は年齢・年次によって異なった様相を見せており、確率推計による不確実性の評価を参考にすることによって、推計結果の活用に対する見方が広がることが示された。また、総人

口のピーク年次の推定についてはその不確実性が高いことを一般的にもわかりやすく示すことができるなど、応用面での活用の可能性も明らかとなり、本研究による手法を利用して、将来人口推計結果を確率的に評価することの有用性が示されたものとする。

謝辞

確率推計に関する研究の機会を与えて頂いた高橋重郷副所長、また、将来人口推計プロジェクトにおいて数々の貴重なコメントを頂いた、金子隆一人口動向研究部長をはじめとするプロジェクトメンバー各位に心より感謝を申し上げたい。特に、守泉理恵氏には「少子化に関する専門家調査」について多大なるご協力を頂いたことを感謝する。また、確率推計についてご教示頂いた稲垣誠一氏、白杵政治氏、北村智紀氏、中嶋邦夫氏にお礼申し上げますとともに、有益なコメントを頂いた査読者にも感謝する。なお、あるべき誤謬などは全て著者の責任である。

参考文献

- 府川哲夫 (2005) 「INAHSIM を用いた世帯の将来推計 (2004)」, 『人口学研究』第36号, pp.1-12.
- Inagaki, S. (2005) "Projections of the Japanese Socioeconomic Structure Using Microsimulation Model (INAHSIM)", *IPSS Discussion Paper Series*, Vol.2005-03.
- 石井太 (2000) 「基礎年金被保険者期間推計へのマイクロシミュレーションの応用」, 『生存科学』第11B巻, pp.23-42.
- 加藤久和 (2005) 「確率的手法に基づく出生率の将来推計」, 『政経論叢』第74巻第1, 2号, pp.265-302.
- 北村智紀, 中嶋邦夫 (2004) 「2004年厚生年金改革案のリスク分析」, 『ニッセイ基礎研所報』第32巻, pp.1-30.
- 国立社会保障・人口問題研究所編 (2002) 『日本の将来推計人口—平成14年1月推計—』, 厚生統計協会.
- 厚生省人口問題研究所編 (1984) 『出生力の生物人口学的分析 昭和55～58年「人口推計の精密化とそのための人口モデルの開発に関する総合的研究の概要報告」』, (特別研究報告資料).
- (1990) 『家族形成モデルの開発と応用に関する研究 平成元～3年度特別研究 第I報告書』, (特別研究報告資料第11号).
- 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, September.
- Lee, R. (1998) "Probabilistic Approaches to Population Forecasting", *Population and Development Review*, Vol.24 No.5, pp.156-190.
- Lutz, W., W. Sanderson, and S. Scherbov (1998) "Expert-Based Probabilistic Population Projections", *Population and Development Review*, Vol.24 No.5, pp.139-155.
- 守泉理恵 (2004) 「少子化に関する専門家調査の分析：専門家による人口をめぐる将来予測と政策評価」, 大淵寛, 高橋重郷編著『少子化の人口学』(人口学ライブラリー1) 原書房, pp.187-211.
- Nelsen, R. B. (2006) *An Introduction to Copulas*: Springer.
- ジェフリー S・シモノフ (1999) 『平滑化とノンパラメトリック回帰への招待』, 農林統計協会.
- 鈴木亘, 湯田道生, 川崎一泰 (2003) 「人口予測の不確実性と年金財政：モンテカルロシミュレーションを用いた人口予測の信頼区間算出と年金財政収支への影響」, 『会計検査研究』, 第28巻, pp.101-112.
- 戸坂凡展, 吉羽要直 (2005) 「コピュラの金融実務での具体的な活用方法の解説」, *IMES DISCUSSION PAPER SERIES*, 第2005-J-19巻, pp.1-49.

Evaluation of the Uncertainties of Population Projections by Probabilistic Approach

Futoshi ISHII

In the official "Population projection for Japan (2002)" by the National Institute of Population and Social Security Research (IPSS), the deterministic approach was used, and the fertility assumptions were based on three scenarios, high, medium and low variant. This is one approach to provide the uncertainties of the population projection. But in this approach, we cannot present the confidence intervals of the results. In this study, we have used a probabilistic approach and evaluated the uncertainties for population projection.

Using probabilistic approach, the distributions of the fertility or mortality assumptions are needed. There are several methods to accomplish that using such as "ex post analysis", "expert opinions", "time series analysis", etc. In this study, we used the distributions of "expert opinions". "The survey about future prospects of low fertility for experts" is performed in 2001. In this survey, the life expectancy at birth in 2050 and the total fertility rates in 2025 are surveyed. These distributions (after smoothing) are used as those of assumptions. The distributions are linearly transformed taking the same mean values as official projection's assumptions at 2050 (life expectancy) or 2025 (TFR). In other points, coefficient variances are set by linear interpolations with 2000 as 0.

It is essential that the temporal correlation of errors for fertility and mortality. In this study, temporal correlations are expressed using normal copulas. Using this method, the distribution in every year is transformed expert opinions' one and has correlation with another year.

Through these processes, we can estimate the confidence interval for total population and the age proportions of the population, and the distribution for the year that gives a peak of the total populations.

The proposed method in this study is proved to be useful since we can use it independently after we have finished with the official deterministic projection. And the results that we have got have shown many advantages of the probabilistic approach.