

特集 I : 世帯推計

世帯規模別分布に関する統計分析

鈴木 透

世帯単位でみた規模別分布と、個人単位でみた所属世帯規模別分布（プロペンシティ）のモメント間の関係について整理する。1950～2015年国勢調査から求めたモメントによると、分散は平均より急速に縮小したが、1975～95年には例外的に平均の縮小の方が急激だった。単独世帯割合は直線的に増加しており、増加率はむしろ減速している。与えられた平均に合わせて世帯規模別分布を変化させる線型変化比モデルを提示し、全国世帯数の将来推計と比較する。また2015年国勢調査の都道府県別平均世帯に適用し、実際の世帯規模別分布と比較する。統計表が「1人以上」でまとめられていることで生じる分散の過小評価を補正する方法を論じる。戦前の国勢調査データに対し、本稿で論じた手法を適用する。

キーワード：世帯規模、モメント、プロペンシティ

I. はじめに

世帯規模の縮小、特に単独世帯の増加は、福祉行政はもちろん住宅需要や消費行動等の幅広い分野に重要な影響をもたらす。国立社会保障・人口問題研究所（2018）の世帯数の将来推計では、将来の平均世帯規模と、「単独」「夫婦のみ」「夫婦と子」「ひとり親と子」「その他」の5類型別世帯数が得られる。しかし世帯規模の分散等の高次モメントや、単独（1人）世帯以外の規模別分布は得られない。本稿は将来の平均世帯規模から、世帯規模別分布を導くリレーショナル・モデルについて論じる。

世帯単位でみた世帯規模別分布に加え、世帯員単位でみた所属世帯規模別分布（プロペンシティ）を考えることができる。世帯推計の手法としてのプロペンシティ法は、世帯員単位の分布こそが個々人の行動を反映した基本的な概念であると考え、将来の所属世帯別分布を仮定した上で、それを世帯別分布に置き換える方法である。世帯規模以外の世帯類型別に将来推計を行う場合、将来の類型別平均規模を合わせて仮定する必要がある。

多くの人口分析では平均等の代表値を論じれば十分だが、それでは不足な場合に分散が用いられることもある。3次モメントが必要とされることはほとんどないが、所属世帯規模分布（プロペンシティ）の分散には世帯規模分布の3次モメントが関わっている。そこで以下ではまず、二種類の規模別分布のモメント間の関係について整理しておく。次いで平均と分散の変化、変化比の直線性を仮定した規模別分布の予測モデルとその適用例について論じる。世帯規模別世帯数の表はスペースを取るのので、附表としてまとめた。

II. 世帯規模分布と所属世帯規模分布

世帯類型 i の世帯数を H_i 、その平均規模を h_i とする。類型 i に所属する世帯員は $M_i = h_i H_i$ 人ということになる。世帯類型別分布と所属世帯類型別分布（プロペンシティ）は、

$$\text{世帯類型分布} \quad f_i = H_i / \sum_i H_i,$$

$$\text{所属世帯類型分布（プロペンシティ）} \quad g_i = M_i / \sum_i M_i = h_i H_i / \sum_i h_i H_i.$$

プロペンシティ法は将来の g_i を設定し、それを別に得られた将来推計人口に乗じて所属世帯類型別人口 M_i を求め、仮定された将来の類型別平均規模 h_i を用いて類型別世帯数 $H_i = M_i / h_i$ を得る方法である。オーストラリアとニュージーランドの公式推計では、世帯主率法ではなくプロペンシティ法が用いられている（Australian Bureau of Statistics 2010, Statistics New Zealand, 2004）。これは世帯主や世帯の代表者という概念を回避し、世帯員をすべて平等に扱うことで、フェミニスト等からの批判を避けるためと思われる。

以下では類型が世帯規模そのもので、 $M_i = i H_i$ となる場合について考える。まず世帯規模別分布 f_i による次のモメントを定義しておく。

$$\text{(平均)} \quad h = \sum_i i f_i,$$

$$\text{(分散)} \quad \sigma_h^2 = \sum_i i^2 f_i - h^2,$$

$$\text{(3次モメント)} \quad \kappa_h^3 = \sum_i i^3 f_i - 3\sigma_h^2 h - h^3.$$

所属世帯規模別分布（プロペンシティ）は、平均世帯規模 h を用いて $g_i = i f_i / h$ となる。所属世帯規模の平均は、平均世帯規模 h に変動係数を加えたものになる。

$$m = h + \sigma_h^2 / h. \quad (1)$$

この関係は娘数（子ども数）と姉妹数（きょうだい数）の関係と同じで、良く知られている（Goodman, et al. 1975, Pullum, 1982）。確率過程でいう時間平均と事象平均の関係も同じ形式である（宮沢 1993）。なお、統計表の最大カテゴリーが「 I 人以上」とまとめられている場合、 σ_h^2 の値は正確ではなく、したがって m も正確な値は得られないことになる。

所属世帯規模の平均に世帯規模の分散が関わっているように、所属世帯規模の分散には世帯規模の3次モメントが関わってくる。

$$\sigma_m^2 = \kappa_h^3/h + \sigma_h^2(1 - \sigma_h^2/h^2). \quad (2)$$

平均については常に $h < m$ が成り立つが、 σ_h^2 と σ_m^2 の大小関係はすぐにはわからない。特別な場合として、 f_i がポアソン分布であれば $h = \sigma_h^2 = \kappa_h^3$ だから、 $\sigma_h^2 = \sigma_m^2$ となる。また幾何分布（失敗の確率 π で i 回目に失敗する）であれば、モメントは次のようになる。

$$h = 1/\pi, \sigma_h^2 = (1-\pi)/\pi^2, \kappa_h^3 = (1-\pi)(2-\pi)/\pi^3.$$

モメント間の関係は $\sigma_h^2 = h(h-1)$ 、 $\kappa_h^3 = \sigma_h^2(2h-1)$ となり、幾何分布では $\sigma_m^2 = 2\sigma_h^2$ となることが確認できる。

表1 世帯規模と所属世帯規模のモメント

	世帯規模分布			所属世帯規模分布		比較	
	h	σ_h^2	κ_h^3	m	σ_m^2	m/h	σ_m^2/σ_h^2
(普通世帯)							
1950	4.9735	5.7808	6.6573	6.1358	5.7684	1.2337	0.9978
1955	4.9697	5.1847	6.1120	6.0130	5.3261	1.2099	1.0273
1960	4.5442	4.4500	4.9639	5.5235	4.5834	1.2155	1.0300
1965	4.0494	3.6165	3.8737	4.9425	3.7755	1.2205	1.0440
1970	3.6883	2.9447	2.4549	4.4867	2.9729	1.2165	1.0096
1975	3.4528	2.6857	1.9797	4.2306	2.6540	1.2253	0.9882
1980	3.3347	2.6312	1.7858	4.1237	2.5441	1.2366	0.9669
1985	3.2302	2.6239	1.9168	4.0425	2.5574	1.2515	0.9747
1990	3.0637	2.5792	2.3064	3.9056	2.6233	1.2748	1.0171
1995	2.8773	2.4541	2.6451	3.7303	2.6460	1.2964	1.0782
2000	2.7126	2.2594	2.7096	3.5455	2.5645	1.3071	1.1350
2005	2.5821	2.0856	2.6385	3.3898	2.4550	1.3128	1.1771
2010	2.4497	1.9303	2.5487	3.2377	2.3498	1.3217	1.2173
2015	2.3553	1.8058	2.4134	3.1220	2.2426	1.3255	1.2419
(一般世帯)							
1980	3.2228	2.7538	2.0162	4.0772	2.6493	1.2651	0.9620
1985	3.1420	2.7090	2.1197	4.0042	2.6403	1.2744	0.9746
1990	2.9885	2.6348	2.4969	3.8702	2.6930	1.2950	1.0221
1995	2.8165	2.4851	2.7987	3.6989	2.7003	1.3133	1.0866
2000	2.6661	2.2756	2.8172	3.5196	2.6037	1.3201	1.1442
2005	2.5472	2.0935	2.7121	3.3691	2.4828	1.3227	1.1859
2010	2.4217	1.9329	2.6034	3.2198	2.3709	1.3296	1.2266
2015	2.3306	1.8057	2.4577	3.1054	2.2600	1.3324	1.2516

国勢調査における世帯規模別世帯数を附表 1 に、それに依拠して計算した日本の普通世帯および一般世帯の各種モメントの推移を表 1 に示した。平均はもちろん縮小しているが、所属世帯規模平均 m は世帯規模平均 h より変化がやや緩慢で、 m/h 比は増加傾向にある。世帯規模分散 σ_h^2 は、1950年および1975~85年には所属世帯規模分散 σ_m^2 より大きかったが、近年は σ_m^2 のの方が大きく、その比 σ_m^2/σ_h^2 は拡大する傾向にある。したがって世帯規模分布は1990年代まではポアソン分布に近い形だったのが、近年は幾何分布に近い形に向かって変化し始めたと考えられることもできよう。3次モメントが常に正なのは、世帯規模分布が h の右に尾を引く形であることを示唆している。

III. 平均に対する弾力性

図 1 には世帯規模の平均 (h) と分散 (σ_h^2) の関係を散布図で示した。変化率がわかりやすいよう縦横軸とも常用対数を用いており、0.5は $\sqrt{10} \approx 3.1623$ に対応する。普通世帯の規模は1950~55年にはほとんど変わらず、分散だけが縮小した。これは図 2 にみるように、単独世帯と3人世帯の割合が低下し、それ以外が増加したことによる。1950年代に50%以上を占めた5人以上世帯は1960~70年代に急減し、4人世帯への集中が進んだ。この時期、分散は平均より急激に縮小した。1975~95年には平均が急速に縮小したのに比べ、分散の変化は緩慢だった。この時期は平均に近い3人世帯の割合が安定しており、単独・2人世帯の割合の上昇と4人以上世帯の割合の低下が、分散より平均を急速に縮小させたためと見られる。1995年以後は分散の縮小が加速したが、これは3人世帯の割合も低下に転じ、2人世帯割合の上昇が減速し、単独世帯への集中が進んだことによる。

図 1 平均と分散の変化

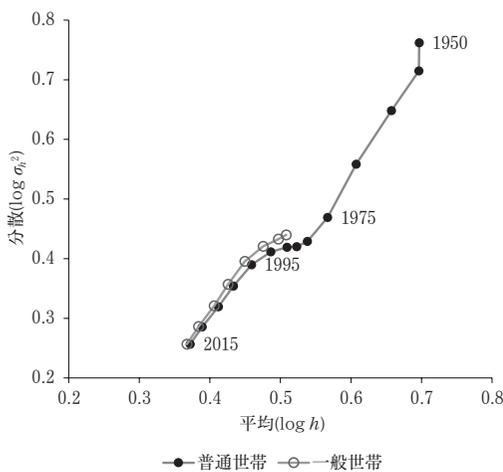
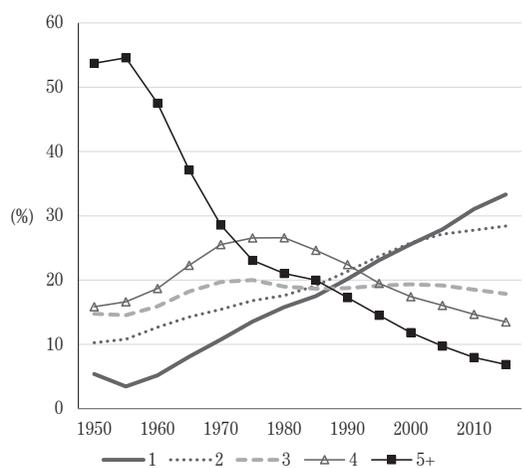


図 2 普通世帯規模別分布(f_i)の推移



このように分散の縮小が平均の縮小より緩慢だった時期もあるが、1950～2015年全体では分散の方が大きく低下しており、弾力性は1より大きいと考えられる。2000～15年の一般世帯で見た場合、平均の年平均増加率は-0.89%、分散は-1.53%で、弾力性は1.7程度になる。つまり平均が1%縮小すれば、分散は1.7%縮小することになる。

図3は平均と単独世帯割合の関係を示したもので、やはり縦横軸とも常用対数で表示しており、-1は0.1（10%）に該当する。これによると単独世帯割合の増加率は1985年以前の方が高く、意外にも近年はむしろ減速したことになる。実際、図2を見ると単独世帯割合は1955年以後ほぼ直線的に増加しており、増加率で見れば減速している。2000～15年の一般世帯で見た場合、平均の年平均増加率は前述のように-0.89%、単独世帯割合は1.51%で、弾力性は-1.7程度になる。つまり平均が1%縮小すれば単独世帯割合は1.7%増加していることになり、符号が異なるだけで分散の弾力性とほぼ同じである。

図3 平均と単独割合の推移

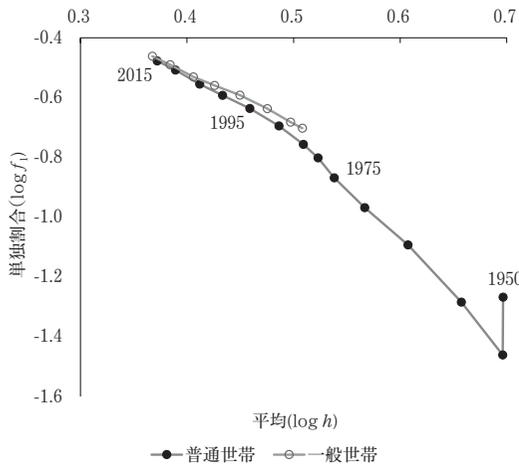
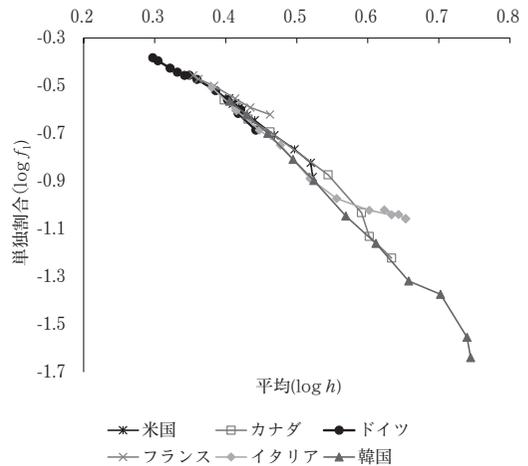


図4 平均と単独割合の推移：諸外国

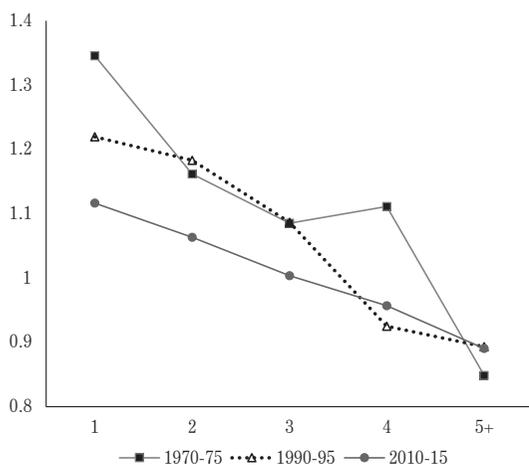


分散と異なり単独世帯割合は比較的入手しやすいので、図4には米国、カナダ、ドイツ、フランス、イタリア、韓国の平均と単独世帯割合の関係を示した（附表2も参照）。これによると各国の世帯変動は、おおむね似通った軌跡を示している。フランスとドイツの弾力性が低いので、変化が進むほど弾力性が低下するように見えるが、国別に見て必ずしも弾力性が低下するわけではない。グラフが重なっていてわかりにくいのが、最も特異な動きをしているのは米国で、1980年以後は平均の縮小も単独世帯割合の上昇も非常に緩慢で、一時的に変化の方向が逆転した期間もある。このため、近年の弾力性はきわめて不安定である。これは先進国としては出生率が高く、同棲カップルを含めればユニオン形成も減少しておらず、離婚率も1970年代以降は低下しているなど（Jacobsen, et al., 2012）、小世帯化・単独世帯化を促進する要因が少ないためと考えられる。それ以外の国の1999年以後データがある最新年次までの弾力性は、フランス（-1.48）は絶対値が日本より小さく、

カナダ (-1.83) とドイツ (-1.75) は日本と同程度で、イタリア (-2.85) と韓国 (-2.38) は日本より大きい。

IV. 線型変化比モデル

図5 普通世帯のプロペンシティの変化比



プロペンシティ法の一例として、与えられた平均世帯規模 h^* に合わせてプロペンシティを設定する方法が考えられる。ここでは2回の国勢調査間のプロペンシティの変化比に $g_i^{(2)}/g_i^{(1)}$ 注目する。図5には1970～75年、1990～95年、2010～15年の変化比を示した。1970～75年は4人世帯への集中が進んでいた時期で、4人世帯の変化比が3人世帯より高い。しかし近年は変化比がおおむね直線上に並ぶようになり、特に2010～15年はそうである。この傾向は、一般世帯で見ても同様である。そこで基準人口におけるプロペンシティ g_i と、仮定された

将来の平均世帯規模 h^* を用いて、次の条件下で g_i^*/g_i を直線で表すことを考える。

$$\sum_i g_i^* = 1, \quad (3a)$$

$$\sum_i \frac{g_i^*}{i} = \frac{1}{h^*}. \quad (3b)$$

結果は次のようになる。

$$\frac{g_i^*}{g_i} = 1 + \frac{m}{m-h} \frac{h-h^*}{h^*} - \frac{m}{m-h} \frac{h-h^*}{h^*} i. \quad (4)$$

これを $f_i = h g_i/i$ および(1)を用いて世帯規模の変化比に直すと、

$$\frac{f_i^*}{f_i} = 1 + \frac{h-h^*}{\sigma_h^2} h - \frac{h-h^*}{\sigma_h^2} i. \quad (5)$$

二つの直線を比較すると、プロペンシティの変化比は世帯規模の変化比を h/h^* 倍しただけであることがわかる。

$$\frac{g_i^*}{g_i} = \frac{h}{h^*} \frac{f_i^*}{f_i}. \quad (6)$$

つまり与えられた平均世帯規模に合わせて世帯規模分布を設定しても所属世帯規模分布（プロペンシティ）を設定しても同じことで、両者は簡単に変換できる。以下では世帯規模分布の変化比に関する直線(5)について考察する。 f_i^* にもとづく分散は、

$$\sigma_h^{2*} = \frac{h^*}{h} \sigma_h^2 + (h-h^*) \left(h^* - h \frac{\sigma_m^2}{\sigma_h^2} \right). \quad (7)$$

右辺第2項に着目すると、平均が縮小していれば第1のカッコは正である。表1で見たとおり1990年以後は $\sigma_m^2/\sigma_h^2 > 1$ だから、第2のカッコは負になるだろう。したがって線型変化比モデルでは、分散の弾力性は1より小さくなると思われる。

単独世帯割合の変化率は、次のように表せる。

$$\frac{f_1^* - f_1}{f_1} = - \frac{h(h-1)}{\sigma_h^2} \frac{h^* - h}{h}. \quad (8)$$

右辺の $(h^* - h)/h$ は平均の変化率だから、 $-h(h-1)/\sigma_h^2$ が弾力性ということになる。近年は $h/\sigma_h^2 > 1$ なので、弾力性の絶対値は1より大きいと考えてよいだろう。

V. 全国推計への適用

表2は国勢調査の規模別一般世帯数分布を基準とし、5～15年後の平均世帯規模 h^* が正確に予測できた場合の規模別分布の予測値の評価である。予測値を f_i^* 、5～15年後の実測値を $f_i^{(c)}$ として、誤差の絶対値の合計は次のように求めた。

$$\sum_{i=1}^5 |f_i^{(c)} - f_i^*|. \quad (9)$$

表2によると、最近ほど予測の精度は向上している。これは図5にみるように最近ほど変化比が直線上に並ぶようになったことに加え、平均の縮小が減速していることも寄与している。平均が大きく変化すれば、それだけ分布の変化も大きく、誤差を生じる余地が増える。また平均の変化が大きければ直線の傾きも大きくなり、負の値が生じる可能性も高まる。一般に $f_i^* < 0$ となる条件は、

$$\frac{\sigma_h^2}{h-h^*} < i-h. \quad (10)$$

表1で見たように2015年の $h = 2.33$ 、 $\sigma_h^2 = 1.8$ だから、 $h-h^*$ が0.674を超えると $5 - 2.33 = 2.67$ を下回る。つまり平均が $2.33 - 0.674 = 1.656$ まで低下しない限り、負の値が生じる心配はしなくてよい。

表2 一般世帯規模分布の予測の精度

期間	パラメタ				予測値					誤差の絶対値の合計
	h	σ_h^2	h^*	$h^* - h$	1	2	3	4	5+	
(5年間)										
1980~1985	3.2228	2.7538	3.1420	-0.0807	0.2113	0.1735	0.1819	0.2474	0.1859	0.1023
1985~1990	3.1420	2.7090	2.9885	-0.1535	0.2331	0.1958	0.1808	0.2251	0.1651	0.0929
1990~1995	2.9885	2.6348	2.8165	-0.1720	0.2608	0.2191	0.1806	0.2018	0.1377	0.0909
1995~2000	2.8165	2.4851	2.6661	-0.1505	0.2842	0.2410	0.1832	0.1750	0.1167	0.0780
2000~2005	2.6661	2.2756	2.5472	-0.1188	0.3000	0.2598	0.1850	0.1576	0.0976	0.0535
2005~2010	2.5472	2.0935	2.4217	-0.1255	0.3220	0.2742	0.1823	0.1434	0.0781	0.0330
2010~2015	2.4217	1.9329	2.3306	-0.0911	0.3454	0.2779	0.1768	0.1332	0.0667	0.0246
(10年間)										
1980~1990	3.2228	2.7538	2.9885	-0.2342	0.2358	0.1849	0.1842	0.2365	0.1586	0.1147
1985~1995	3.1420	2.7090	2.8165	-0.3255	0.2614	0.2092	0.1825	0.2123	0.1348	0.1070
1990~2000	2.9885	2.6348	2.6661	-0.3225	0.2871	0.2307	0.1805	0.1893	0.1124	0.1039
1995~2005	2.8165	2.4851	2.5472	-0.2693	0.3064	0.2499	0.1815	0.1644	0.0977	0.0801
2000~2010	2.6661	2.2756	2.4217	-0.2444	0.3254	0.2690	0.1816	0.1451	0.0790	0.0358
2005~2015	2.5472	2.0935	2.3306	-0.2166	0.3418	0.2805	0.1787	0.1335	0.0656	0.0295
(15年間)										
1980~1995	3.2228	2.7538	2.8165	-0.4062	0.2634	0.1977	0.1867	0.2242	0.1280	0.1244
1985~2000	3.1420	2.7090	2.6661	-0.4760	0.2861	0.2208	0.1839	0.2010	0.1082	0.1169
1990~2005	2.9885	2.6348	2.5472	-0.4413	0.3078	0.2399	0.1804	0.1795	0.0925	0.1026
1995~2010	2.8165	2.4851	2.4217	-0.3949	0.3299	0.2594	0.1798	0.1531	0.0778	0.0585
2000~2015	2.6661	2.2756	2.3306	-0.3354	0.3438	0.2757	0.1791	0.1361	0.0654	0.0322

表3は将来の規模別分布の予測値である。基本パラメタである平均 h と分散 σ_h^2 は2015年の値、将来の平均 h^* は国立社会保障・人口問題研究所(2018)の将来推計値を用い、線型変化比モデルによって将来の規模別分布を求めた。それを一般世帯数の将来推計値に乘じて規模別世帯数を求めた。将来推計では規模別世帯数ではなく家族類型(「単独」「夫婦のみ」「夫婦と子」「ひとり親と子」「その他」)別の一般世帯数を求めており、直接比較できるのは単独世帯数だけである。表3によると、線型変化比モデルは将来推計より単独世帯数を多めに評価するが、その差は小さいと言える。

表3 規模別一般世帯数の予測値と単独世帯数の評価

	平均(h)	一般世帯数(千世帯)						単独世帯 ¹⁾	誤差率(%)
		総数 ¹⁾	1	2	3	4	5+		
2015	2.33	53,332	18,418	14,877	9,365	7,069	3,603	18,418	—
2020	2.26	54,107	19,722	15,301	9,236	6,673	3,175	19,342	1.96
2025	2.19	54,116	20,576	15,474	9,020	6,265	2,782	19,960	3.09
2030	2.15	53,484	20,992	15,425	8,747	5,875	2,445	20,254	3.64
2035	2.10	52,315	21,019	15,185	8,431	5,513	2,167	20,233	3.88
2040	2.08	50,757	20,735	14,802	8,093	5,184	1,944	19,944	3.96

1) 国立社会保障・人口問題研究所(2018)による。

VI. 都道府県別平均への適用

2015年国勢調査の全国における世帯規模別分布 f_i とそのモメント h および σ_h^2 を基準とし、都道府県別平均世帯規模 h^* を用いて世帯規模分布 f_i^* を予測し、それを国勢調査の実測値と比較した結果を表4に示した。

表4 都道府県別世帯規模分布の予測の評価：2015年

	平均	分散	誤差					誤差の絶対
	h	σ_h^2	$f_1^* - f_1$	$f_2^* - f_2$	$f_3^* - f_3$	$f_4^* - f_4$	$f_{5+}^* - f_{5+}$	値の和
全国	2.3306	1.8057	—	—	—	—	—	—
北海道	2.1320	1.4118	0.0230	-0.0318	-0.0006	0.0057	0.0037	0.0648
青森県	2.4820	2.0967	0.0055	-0.0196	-0.0023	0.0248	-0.0085	0.0606
岩手県	2.5381	2.3213	-0.0110	-0.0089	0.0053	0.0304	-0.0157	0.0713
宮城県	2.4271	2.1433	-0.0228	0.0180	0.0043	0.0109	-0.0104	0.0665
秋田県	2.5538	2.1764	0.0094	-0.0284	-0.0010	0.0296	-0.0097	0.0780
山形県	2.7847	2.6508	-0.0252	-0.0050	0.0147	0.0421	-0.0267	0.1136
福島県	2.5569	2.3763	-0.0182	-0.0019	0.0098	0.0276	-0.0173	0.0748
茨城県	2.5462	2.0796	0.0069	-0.0144	-0.0017	0.0119	-0.0027	0.0375
栃木県	2.5400	2.0679	0.0037	-0.0077	-0.0034	0.0100	-0.0025	0.0272
群馬県	2.4957	1.8956	0.0170	-0.0175	-0.0064	0.0031	0.0038	0.0478
埼玉県	2.4081	1.6930	0.0208	-0.0096	-0.0146	-0.0102	0.0135	0.0687
千葉県	2.3488	1.6789	0.0171	-0.0108	-0.0090	-0.0083	0.0110	0.0561
東京都	1.9901	1.3959	-0.0410	0.0545	0.0053	-0.0141	-0.0048	0.1197
神奈川県	2.2612	1.5587	0.0080	0.0059	-0.0110	-0.0153	0.0123	0.0525
新潟県	2.6536	2.3935	-0.0128	-0.0062	0.0069	0.0292	-0.0171	0.0723
富山県	2.6640	2.2603	-0.0010	-0.0116	0.0012	0.0213	-0.0100	0.0451
石川県	2.4756	2.0611	-0.0066	0.0005	0.0054	0.0079	-0.0071	0.0275
福井県	2.7492	2.5541	-0.0250	-0.0008	0.0192	0.0318	-0.0251	0.1020
山梨県	2.4673	1.8993	0.0153	-0.0200	-0.0031	0.0079	-0.0001	0.0464
長野県	2.5464	2.0802	0.0118	-0.0261	0.0025	0.0182	-0.0065	0.0651
岐阜県	2.6499	2.1745	0.0061	-0.0185	0.0061	0.0124	-0.0061	0.0491
静岡県	2.5394	2.0580	0.0069	-0.0142	-0.0010	0.0121	-0.0039	0.0381
愛知県	2.4075	1.9098	-0.0090	0.0139	0.0028	-0.0078	0.0001	0.0336
三重県	2.4715	1.8963	0.0153	-0.0206	0.0021	0.0006	0.0026	0.0411
滋賀県	2.5928	2.1221	-0.0059	0.0036	0.0068	-0.0009	-0.0035	0.0208
京都府	2.2220	1.6575	-0.0092	0.0135	0.0038	-0.0082	0.0001	0.0347
大阪府	2.2172	1.5859	-0.0011	0.0066	0.0018	-0.0113	0.0040	0.0249
兵庫県	2.3491	1.7115	0.0136	-0.0110	-0.0029	-0.0068	0.0071	0.0414
奈良県	2.5225	1.7759	0.0395	-0.0416	-0.0081	-0.0030	0.0132	0.1054
和歌山県	2.3995	1.7177	0.0343	-0.0419	-0.0040	0.0046	0.0070	0.0918
鳥取県	2.5675	2.2940	-0.0098	-0.0084	0.0079	0.0264	-0.0160	0.0685
島根県	2.5344	2.3215	-0.0086	-0.0177	0.0115	0.0345	-0.0198	0.0921
岡山県	2.4259	1.9691	-0.0011	-0.0053	0.0056	0.0057	-0.0049	0.0226
広島県	2.2937	1.6792	0.0098	-0.0100	-0.0006	-0.0024	0.0031	0.0258
山口県	2.2704	1.6106	0.0274	-0.0390	0.0015	0.0083	0.0018	0.0781
徳島県	2.3927	1.8869	0.0080	-0.0147	0.0006	0.0078	-0.0016	0.0327
香川県	2.3862	1.8114	0.0157	-0.0212	-0.0006	0.0044	0.0018	0.0436
愛媛県	2.2807	1.6280	0.0222	-0.0283	-0.0019	0.0048	0.0032	0.0604
高知県	2.2021	1.5762	0.0138	-0.0213	0.0008	0.0073	-0.0006	0.0437
福岡県	2.2581	1.7538	-0.0099	0.0105	0.0037	-0.0004	-0.0040	0.0284
佐賀県	2.6743	2.4361	-0.0108	-0.0129	0.0105	0.0321	-0.0189	0.0852
長崎県	2.3716	1.8762	0.0156	-0.0306	0.0026	0.0175	-0.0050	0.0713
熊本県	2.4596	2.0606	0.0033	-0.0174	0.0053	0.0165	-0.0078	0.0502
大分県	2.3207	1.7577	0.0158	-0.0261	0.0025	0.0095	-0.0017	0.0556
宮崎県	2.3104	1.6773	0.0293	-0.0441	0.0027	0.0130	-0.0009	0.0899
鹿児島県	2.1972	1.5202	0.0227	-0.0359	0.0070	0.0082	-0.0020	0.0757
沖縄県	2.5012	2.1835	-0.0217	0.0158	0.0071	0.0179	-0.0191	0.0816

予測値との乖離が最も大きいのは東京都、次いで山形県で、これらは平均世帯規模が全国値から大きく離れている都県である。3番目に乖離が大きい奈良県は、全国平均世帯規模との差は0.1919で、東京都(-0.3406)や山形県(0.4541)よりはるかに小さい。実際、平均の差の絶対値では第19位に過ぎない。また東京都や山形県では、単独世帯が理論値より多いのに対し、奈良県は2~4人世帯が理論値より多く、分散が小さい分布になっている。

図6に平均(h)と分散(σ_h^2)の散布図を示した。回帰係数が弾力性を表すよう、双方とも対数をとった。確かに奈良県の分散は平均に対し小さく、全体の傾向から乖離している。一方で図7に見るように東京都は単独世帯が異様に多いため、平均に比べ分散が大きい。山形県の平均と分散の関係は、全体的な傾向から大きく外れていない。

図6 都道府県別世帯規模の平均と分散

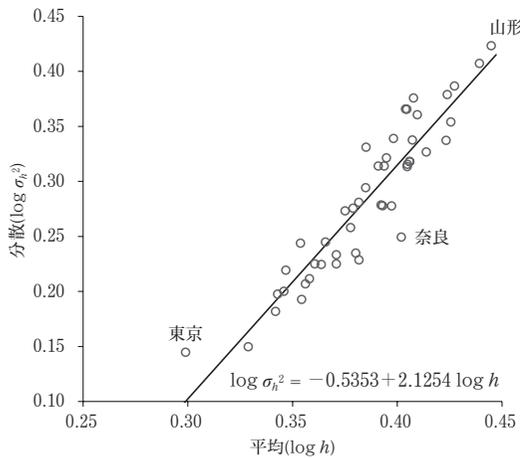
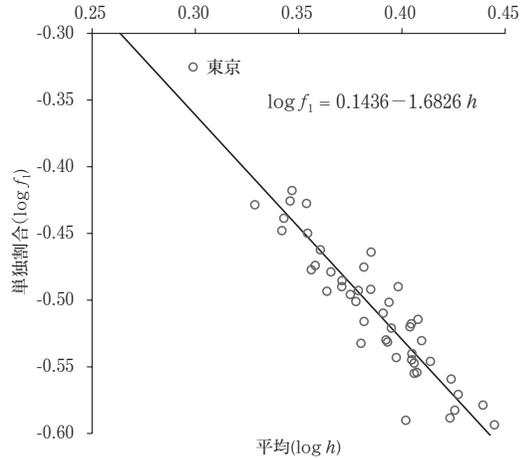


図7 都道府県別世帯規模の平均と単独世帯割合



分散の平均に対する回帰係数は2.1254で、前述した2000~15年の弾力性1.7より大きい。一方で単独世帯割合の回帰係数は-1.6826で、前述の-1.7とほぼ等しい。地域間偏差は時系列的变化に比べ、世帯規模が小さい地域は2~3人世帯に集中する傾向が強いようである。

VII. 分散の過小評価

前述のように実際の国勢調査の統計表は、最後のカテゴリーが「I人以上の世帯」とオープンカテゴリーになっており、このような表にもとづいて計算した分散は過小評価になる。このI人以上の世帯の規模の条件付平均を h_I とする。実際の統計表では、 h_I はI人以上の世帯数/I人以上の世帯に所属する世帯員数で求められる。I人未満の世帯については、i人世帯の世帯数×iで所属成員数が得られるから、I-1人世帯までの合計を世帯員総数

から引けば I 人以上の世帯に所属する世帯員数が得られる。統計表から計算される分散は、

$$\sigma_h^2 = \sum_{i=1}^{I-1} i^2 f_i + h_I^2(1 - F_{I-1}) - h_2. \quad (11)$$

$I+1$ 人以上の世帯が 1 世帯でもあれば、(11)で計算される分散は真の分散より小さくなるだろう。過小評価の度合いは、 I 人以上の世帯の割合 $1 - F_{I-1}$ が大きいほど大きくなることが予想される。2010年国勢調査では、普通世帯でも一般世帯でも $1 - F_{I-1}$ は 0.014% で、過小評価があってもごくわずかだろう。しかし1950年の普通世帯では、 $1 - F_{I-1}$ は 4% を超えており、計算された分散はかなりの過小評価になっている可能性がある。

ここでは I 人以上の世帯が失敗確率 $\pi = 1/(h_I - I + 1)$ の幾何分布に従うと仮定し、分散の過小評価を補正する方法を考える。確率関数は、

$$f_{I-1+j} = (1 - F_{I-1})(1 - \pi)^{j-1}\pi, \quad j = 1, \dots, \infty. \quad (12)$$

この場合、正しい分散は(11)で計算される分散に次のような補正項が付いた形になる。

$$\sigma_h^{2*} = \sigma_h^2 + (1 - F_I)(h_I - I + 1)(h_I - I). \quad (13)$$

補正項は I 人以上の世帯の割合 $1 - F_{I-1}$ に比例して大きくなり、条件付平均 h_I に対しては放物線を描いて大きくなる。

表 5 普通世帯の分散の補正

	10人以上割合 $1 - F_9$	条件付平均 h_{10}	分散 σ_h^2	補正後分散 σ_h^{2*}	差 $\sigma_h^2 - \sigma_h^{2*}$
1950	0.04214	10.82	5.7808	5.8438	0.0630
1955	0.03450	10.90	5.1847	5.2432	0.0585
1960	0.01858	10.71	4.4500	4.4726	0.0226
1965	0.00752	10.79	3.6165	3.6271	0.0106
1970	0.00252	10.58	2.9447	2.9470	0.0023
1975	0.00112	10.57	2.6857	2.6867	0.0010
1980	0.00067	10.63	2.6312	2.6318	0.0007
1985	0.00047	10.75	2.6239	2.6245	0.0006
1990	0.00031	10.67	2.5792	2.5796	0.0003
1995	0.00023	10.59	2.4541	2.4543	0.0002
2000	0.00020	10.54	2.2594	2.2596	0.0002
2005	0.00020	10.52	2.0856	2.0857	0.0002
2010	0.00019	10.56	1.9303	1.9305	0.0002
2015	0.00014	10.57	1.8058	1.8059	0.0001

表5にはこの方法で分散を補正した結果を示した。「10人以上」で丸めたことによる分散の過小評価は、1950年代にはかなり大きく、小数点以下第1位に影響するほどだった。このため1950～60年代の分散の縮小は、図1に示したものよりさらに大きかったと考えられる。しかし10人以上世帯の割合が急減したことにより、近年はほとんど問題にならない程度に縮小した。10人以上世帯の規模の条件付平均も1960年代までは大きく、分散の過小評価を助長していた。条件付平均は1980年代に一時上昇したが、2000年以後は10.52～10.57の範囲で推移している。

VIII. 戦前の国勢調査データの分析

戦前の国勢調査における規模別普通世帯数の統計を見ると、1920年と1930年の報告書では11人以上は5人ごとにまとめて表示し、最終カテゴリーは51人以上の世帯となっている。1930年の朝鮮国勢調査報告でも同様である。実際の統計表は、次のようになっている。

表6 戦前の国勢調査における規模別普通世帯数および世帯人員

規模	日本(1920)			日本(1930)			朝鮮(1930)		
	世帯数	世帯人員	平均規模	世帯数	世帯人員	平均規模	世帯数	世帯人員	平均規模
計	11,122,120	54,336,356	4.89	12,600,276	62,760,821	4.98	3,957,111	20,814,103	5.26
1	641,860	641,860	1	694,063	694,063	1	103,185	103,185	1
2	1,392,026	2,784,052	2	1,480,773	2,961,546	2	342,225	684,450	2
3	1,690,534	5,071,602	3	1,870,115	5,610,345	3	557,497	1,672,491	3
4	1,698,893	6,795,572	4	1,905,489	7,621,956	4	679,805	2,719,220	4
5	1,620,484	8,102,420	5	1,826,367	9,131,835	5	674,825	3,374,125	5
6	1,397,347	8,384,082	6	1,596,536	9,579,216	6	553,802	3,322,812	6
7	1,059,924	7,419,468	7	1,243,343	8,703,401	7	391,781	2,742,467	7
8	702,613	5,620,904	8	851,617	6,812,936	8	255,178	2,041,424	8
9	418,650	3,767,850	9	516,311	4,646,799	9	160,943	1,448,487	9
10	240,002	2,400,020	10	297,722	2,977,220	10	108,092	1,080,920	10
11～15	231,334	2,777,478	12.01	287,836	3,446,495	11.97	120,310	1,454,769	12.09
16～20	20,765	360,680	17.37	23,576	408,915	17.34	8,485	145,320	17.13
21～25	4,488	101,512	22.62	4,352	98,190	22.56	710	16,024	22.57
26～30	1,685	46,628	27.67	1,436	39,616	27.59	180	4,970	27.61
31～35	653	21,430	32.82	397	12,993	32.73	56	1,826	32.61
36～40	359	13,591	37.86	170	6,423	37.78	20	748	37.40
41～45	176	7,528	42.77	71	3,032	42.70	5	216	43.20
46～50	119	5,703	47.92	41	1,960	47.80	6	290	48.33
51+	208	13,976	67.19	61	3,880	63.61	6	359	59.83

この場合の分散は、11人以上の世帯については階級の平均規模の二乗を用いて計算することになり、かなりの過小評価が生じると思われる。51人以上世帯については、1920年日本でもその割合は0.0019%に過ぎず、オープンエンドであることによる過小評価はごく小さいだろう。やはり問題は、11～50人世帯に関する情報が不完全なことによる分散の過小

評価がどの程度かである。

一方、1920年台湾国勢調査では、存在するすべての世帯規模について世帯数が示されている。実際の統計表は附表6の通りで、最大世帯規模はなんと149人である。この表の情報を全て利用して計算した正しい分散が12.93だったのに対し、表6と同じ形式にして5人階級ごとの平均規模によって計算した分散は12.78で、1.17%の過小評価になった。

表6の日本と朝鮮の5人階級を、1人ずつに分解することを考える。一般に*i*人～*i*+4人世帯の合計を H_{5i} 、この階級の平均世帯規模を h_{5i} と表記する。階級内の世帯数 $H_i, H_{i+1}, H_{i+2}, H_{i+3}, H_{i+4}$ が公比 R の等比数列に従うと仮定すると、

$$H_{5i} = H_i \frac{1 - R^5}{1 - R}, \quad (14)$$

$$h_{5i} = \frac{H_i}{H_{5i}} \{(i+4)R^4 + (i+3)R^3 + (i+2)R^2 + (i+1)R + i\}. \quad (15)$$

$Q = h_{5i} - i$ として、これらを満たす R は次の4次方程式の根ということになる。

$$(Q-4)R^4 + (Q-3)R^3 + (Q-2)R^2 + (Q-1)R + Q = 0. \quad (16)$$

これを代数的に解くのは難しいので、表7にはニュートン法によって求めた R の値を示した。世帯規模が35人までは、規模が大きくなるほど公比は1に近づき、勾配は小さくなる。36人以上では結果は不安定で、特に1930年の朝鮮では公比が1を超えており、5人階級内で規模の大きい世帯の方がむしろ多いという結果になっている。これは表7に見るように、階級内の総世帯数が5世帯あるいは6世帯と少なく、世帯数が等比数列に従うという仮定が意味をなさないことによる。5世帯を*i*人～*i*+4人世帯に分配する場合の数はたかだか126通り、6世帯なら230通りなので、すべての場合について階級内平均 h_{5i} を計算すれば実際の分配パターンが見出せるだろうが、ここでは行っていない。また分解した結果を整数化することもしていない。

表8には戦前の日本と朝鮮の世帯規模の平均と分散を示した。やはり統計表通りの分布と階級別平均を用いて計算した分散は過小評価になっており、場合によっては小数点以下第1位から変わって来る。分散の過小評価に従い、所属世帯規模の平均も過小評価されることになる。

表7 規模別世帯数推計のための公比

	日本(1920)	日本(1930)	朝鮮(1930)
11～15人	0.5704	0.5569	0.6055
16～20人	0.7191	0.7088	0.6197
21～25人	0.8238	0.7996	0.8025
26～30人	0.8473	0.8106	0.8206
31～35人	0.9126	0.8719	0.8189
36～40人	0.9313	0.8964	0.7316
41～45人	0.8920	0.8613	1.1057
46～50人	0.9629	0.9067	1.1838

表 8 戦前の普通世帯の平均と分散

		日本(1920)	日本(1930)	朝鮮(1930)	
世帯規模の平均	h	4.8854	4.9809	5.2599	
所属世帯規模の平均	(階級平均による)	m	6.2608	6.3240	6.4504
	(補間分布による)	m^*	6.2676	6.3310	6.4596
世帯規模の分散	(階級平均による)	σ_h^2	6.7192	6.6897	6.2618
	(補間分布による)	σ_h^{2*}	6.7525	6.7245	6.3105

戦前の国勢調査で規模別普通世帯数の詳細な統計表が得られたのは、日本は1920年と1930年、朝鮮は1930年のみ、台湾は1920年のみである。一方、表9に示した年次については普通世帯と人員の総数が得られ、正確な平均世帯規模(h)を計算できた。この時期、日本帝国内では平均世帯規模はおおむね上昇傾向にあった。出生率上昇と乳幼児死亡率低下による同居子数の増加が主な原因だろう。

平均のみ得られて規模別分布は得ら

れない年次については、線形変化比モデルで推計し、結果を附表7に示した。日本の1925年は1920年を、1935年と1940年は1930年の分布を出発点とした。日本と朝鮮の基準年における分散は、5人階級内の世帯数を補間し過小評価を補正した値(表8の σ_h^{2*})を用いた。

この期間の最頻値は、日本・朝鮮・台湾いずれも4人世帯で変わっていない。戦後日本では核家族化とともに4人世帯への集中が進み、分散が急速に縮小したが、同じような変化は韓国・台湾でも起きただろう。表10にみるように、単独世帯割合は日本に比べ朝鮮では少なく、台湾では多かった。前述のように1920年の台湾の世帯規模の分散は12.93で、非常に分散が大きい分布をしていた。朝鮮と台湾では平均世帯規模が単調に増加したため、単独世帯割合は減少し11人以上世帯の割合は増加した。しかし総世帯数が増加していたため、単独世帯を含め絶対数が減少する場合は稀だった。推計結果が正しい場合、減少したのは日本の2人世帯(1920~25年)と台湾の単独世帯(1920~25年)だけである。

表 9 戦前の国勢調査における普通世帯の平均規模

	年次	普通世帯数	普通世帯人員	平均規模
日本	1920	11,122,120	54,336,356	4.8854
	1925	11,902,593	58,015,326	4.8742
	1930	12,600,276	62,760,821	4.9809
	1935	13,383,349	67,249,793	5.0249
	1940	14,091,157	70,393,324	4.9956
朝鮮	1925	3,686,691	19,325,461	5.2420
	1930	3,957,111	20,814,103	5.2599
	1935	4,264,332	22,609,232	5.3019
台湾	1920	683,723	3,624,403	5.3010
	1925	726,526	3,993,408	5.4966
	1935	897,565	5,212,426	5.8073
	1940	1,038,883	6,077,478	5.8500

表10 戦前の国勢調査にみる単独世帯の変化

	年次	単独世帯数	割合(%)
日本	1920年	641,860	5.8
	1925年*	691,350	5.8
	1930年	694,063	5.5
	1935年*	718,005	5.4
	1940年*	769,450	5.5
朝鮮	1925年*	97,300	2.6
	1930年	103,185	2.6
	1935年*	108,042	2.5
台湾	1920年	62,864	9.2
	1925年*	62,452	8.6
	1935年*	68,624	7.6
	1940年*	78,071	7.5

*線形変化比モデルによる推計値

IX. おわりに

本稿で使用した線型変化比モデルは汎用的とは言えず、1970年代のように4人世帯への集中が進んでいた時期には精度が悪くなる。戦前のデータには検証なしに適用したが、実際には日本の1920～30年の変化比は直線的とは言い難く、2人世帯の減少が顕著に大きい。朝鮮・台湾については一時点の規模別分布しか得られなかったので、検証のしようがない。今後はさらにデータを収集して適用可能性を見定め、変化比が直線上に並ばない場合の対処法も考える必要がある。

付記

本研究は、厚生労働行政推進調査事業費補助金（政策科学総合研究事業（政策科学推進研究事業））「国際的・地域的視野から見た少子化・高齢化の新潮流に対応した人口分析・将来推計とその応用に関する研究（H29-政策-指定-003）」による助成を受けた。

参考文献

- 国立社会保障・人口問題研究所（2018）『日本の世帯数の将来推計（全国推計）—2015（平成27）～2040（平成52年）— 2018（平成）年推計』人口問題研究資料第339号。
- 宮沢政清『確率と確率過程』近代科学社, 1993.
- Australian Bureau of Statistics (2010) *Household and Family Projections, Australia, 2006 to 2031 Quality Declaration*.
- Goodman, Leo A., Nathan Keyfitz and Thomas W. Pullum (1975) "Addendum to Family Formation and the Frequency of Various Kinship Relationships," *Theoretical Population Biology* 8:376-381.
- Jacobsen, Linda A., Mark Mather and Genevieve Dupuis (2012) "Household Change in the United States," *Population Bulletin* 67(1):2-13.
- Pullum, Thomas W. (1982) "The Eventual Frequencies of Kin in a Stable Population," *Demography* 19(4): 549-565.
- Statistics New Zealand, *New Zealand Family and Household Projections 2001(base)-2021*.

Statistical Analysis of Household Size

SUZUKI Toru

This paper discusses two kinds of household size distributions; one is based on the number of households by size and the other is based on the number of individuals living in household by size. The latter is called "propensity." It is shown that a moment of propensity is a function of a higher moment of household size distribution. An analysis of census data in Japan shows that the change in variance is usually greater than that of mean, with an exception of 1975–95 period when the mean reduced more rapidly than the variance.

A linear ratio of change model is proposed to obtain the distribution of household size according to a given average. This model is applied to the national average in the future and to prefectural averages in the 2015 census. A method to adjust an underestimation of variance is also proposed when a census table does not show all the household sizes but the last category summarize such as "ten persons and more." This method is applied to the post-war censuses of Japan and pre-war censuses of Japan, Korea, and Taiwan to evaluate the extent of underestimation of variances. It is shown that the average household size increased in the Japanese empire in 1920–40, and that the percentage of one-person household was lowest in Korea and highest in Taiwan.

keywords : household size, moments, propensity