

人口集団における先頭数字頻度分布

坂井博通

I はじめに

人口に関する諸現象には、種々の曲線があてはまることが経験的、理論的に知られる。たとえば、人口増加には、修正指数曲線、ゴンパーツ曲線、そしてロジスティック曲線があてはまることが知られている¹⁾。初婚年齢の頻度分布には、Coale and McNeilの分布²⁾、婚約期間、第1妊娠待ち時間にはワイブル分布³⁾、離婚した夫婦の前同居期間もワイブル分布⁴⁾があてはまると言われる。

人口自体の分布に関しては、たとえば、都市人口やその密度分布は対数正規曲線があてはまると言われる⁵⁾。また、都市の規模とその順位に関するジップの法則も知られている。

なぜそれら曲線があてはまるのかは必ずしも十分にわかっていないが、当てはめた結果は補間（外）や推計に用いられ、また、人口現象の表現方法あるいはメカニズムの理解にも有用であると考えられている。

本稿は、種々の人口分布における先頭数字の頻度分布において以下にのべるような規則性を発見したため、それを報告し、その成立機序の若干の考察を行うものである。

II 人口の先頭数字頻度漸減の法則

表1は、世界の国の人口（1988年）の最上位桁、2位の桁、3位の桁に現れる数字の頻度分布を示したものである。第2位と第3位の数字頻度分布は、均等分布の帰無仮説を否定されないが、最上位に関しては、大きく数字の均等分布を離れる（ $P < 0.01$ ）。実際、1で始まる国は52か国で31%も占める。そして、若干のどこぼこはあるものの、数字が大きくなるにつれてその数字で始まる国は少なくなっていく。9で始まる人口はわずか3か国しかない。

表2は、1950、1960、1970年における世界各国の人口の先頭数字の分布を見たものである。その結果、1988年ほどクリアではないが、同様な分布が見られる。したがって、先頭数字が大きくなるにつれて頻度が小さくなるということは、1988年の特殊事情ではないことがわかる。

しかし、この傾向は、世界に1,000万台の国が非常に多いために、全体としてその効果が大きく出ているとも考えられる。

1) 山口喜一編著、『人口分析入門』, 1989年, pp.59-61.

2) Coale A. J. and McNeil D. R., The distribution by Age of the Frequency of First Marriage in a Female Cohort, *Journal of American Statistical Association*, Vol.67, No.4, 1972, pp.743-749.

3) 大谷憲司, 「結婚と妊娠にいたる過程に含まれるいくつかの時間分布について」, 『人口問題研究』, 第45巻第4号, 1991年, pp.1-16.

4) 印東太郎, 『数理科学』, 1974年2月.

5) 鈴木啓祐, 『人口分布の構造解析』, 大明堂, 1985年, p.12.

表1 世界各国の人口の位別数字頻度分布

数 字	最 上 位	2 位	3 位
	(158)	(158)	(64)
0	/	20	7
1	52	19	2
2	26	13	7
3	23	15	2
4	12	20	9
5	14	15	8
6	8	11	7
7	12	15	6
8	8	16	6
9	3	14	10
(カイ2乗値)	100.0	4.1	9.7
	P<0.01	N. S.	N. S.

資料) 国連統計月報1990年3月

注1) 最上位と2位は10万以上の国, 3位は100万以上の国について.

注2) カイ2乗値は, 均等分布を期待値とした差を示す.

よって, 桁数別に見てみた(図1).
そうすると, 1,000万台の国は確かに非常に多く, その全体に与える影響は大きい. しかし, 10-90万台の国, 100-900万台の国においても1で始まる国が多く, そして, 2, 3……9となるにつれて減少していく傾向が見てとれる. したがって, 1000万台の国が多いことによる効果だけではない, ことがわかった.

よって, 各国人口の先頭数字は1がもっとも多く, 以下その頻度が頻度していくと言える(先頭数字頻度漸減の法則と名付ける).

次に, 国別人口分布以外の人口分布に関してこの法則が成立するかをいくつかの例で検討する. 図2は, 世界の人口10万以上の首都の先頭数字の頻度分布を見たものである. 国別人口と同様な分布が観察される.

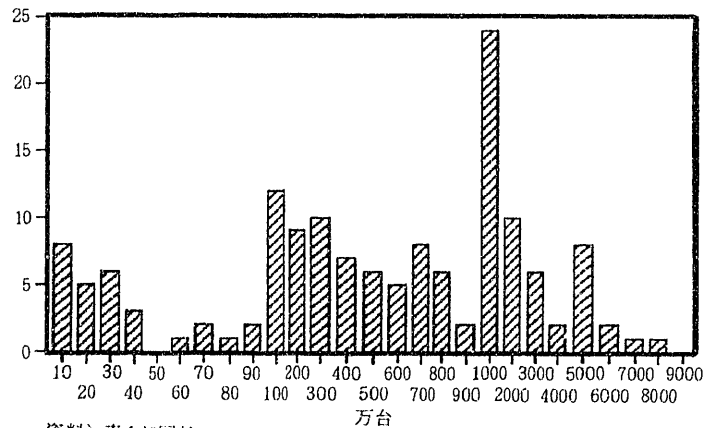
日本の過去から現在の人口に関しては, 2で始まる時期がないものの1で始まる頻度をもっとも多い. また, 1-4のそれぞれの頻度は5-9のそれぞれの頻度よりも大きい. 日本における都道府県別人口に関しては, 3以下の先頭数字は, 必ずしも漸減していく様子が見えな

表2 世界各国の人口の先頭数字頻度分布

数 字	年 次			
	1988	1950	1960	1970
	(158)	(138)	(141)	(142)
1	52	41	34	33
2	26	29	30	32
3	23	13	21	19
4	12	17	15	13
5	14	8	11	17
6	8	4	7	10
7	12	12	7	4
8	8	11	6	8
9	3	3	10	6

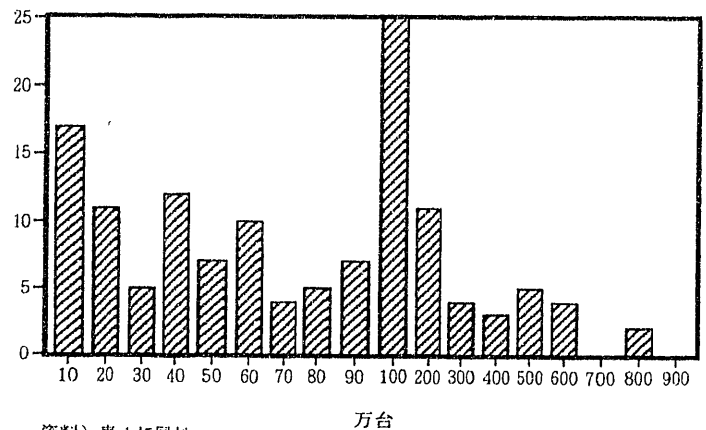
資料) 表1の資料, ならびに世界人口年鑑1978年(上)

図1 世界各国の人口の頻度分布 (10万以上1億未満)



資料) 表1と同じ.

図2 10万以上の首都人口の頻度分布



資料) 表1と同じ.

いが、1が23県と50%近くを占め、続いて8が17%を占める。東京都の場合は、先頭数字が小さい方が頻度が高いということがうかがわれる(表3)。

以上のように、いくつかの地理的人口において、先頭数字頻度漸減の法則が観察された。

Ⅲ 考察

ここでは、先頭数字頻度漸減の法則が成り立つ理由を考えてみる。

いま、1つの集団(人口1,000)を考え、その集団が一定の年間増加率 r で成長すると仮定する。人口が2,000になるまでにかかる時間 t は、次のようになる。

$$1000 \times (1 + r)^t = 2000 \text{ から,}$$

$$t = \log(2000/1000) / \log(1 + r)$$

同様に、2000から3000に要する時間、3000から4000になる時間を求めると表4のようになる。もちろん、桁の異なる人口に関しても同様の計算ができる。そして、先頭数字の期待頻度の比は所要時間の比になると考えられる。

表4 1集団(初期人口1,000)における増加率一定の場合の先頭数字変化の所要時間とその比(全体で100)

先頭数字	人口変化	所要時間	所要時間の比
1	1000 → 2000	$\log(2000/1000) / \log(1+r)$	30.1
2	2000 → 3000	$\log(3000/2000) / \log(1+r)$	17.6
3	3000 → 4000	$\log(4000/3000) / \log(1+r)$	12.4
4	4000 → 5000	$\log(5000/4000) / \log(1+r)$	9.7
5	5000 → 6000	$\log(6000/5000) / \log(1+r)$	7.9
6	6000 → 7000	$\log(7000/6000) / \log(1+r)$	6.7
7	7000 → 8000	$\log(8000/7000) / \log(1+r)$	5.8
8	8000 → 9000	$\log(9000/8000) / \log(1+r)$	5.1
9	9000 → 10000	$\log(10000/9000) / \log(1+r)$	4.6

もし、別の異なる集団がやはり一定の(異なる)増加率で成長をしているとすると、先頭数字に関して表4と同様の分布を期待できる。そして、人口変動が相互に独立であると考え、2つの集団をあわせた先頭数字の期待頻度の分布も表4と同様になる。このように考えると、一般的に n 個の集団の先頭数字の期待頻度の分布は、表4の所要時間の比と同じになると考えられる⁶⁾。

実際、表5は、1から1,000の間の初期人口をもつ1,000の集団に一定の年増加率(0%から10%まで)を与え続けた100年後の結果であるが、全体では表4の所要時間の比とほぼ同じになり、また、5-7桁においても、先頭数字の頻度は漸減していくことがわかる。

6) ちなみに、 $\log(2/1) = \log((10/9) \times (9/8) \times (8/7) \times (7/6) \times (6/5)) = \log(10/9) + \log(9/8) + \log(8/7) + \log(7/6) + \log(6/5)$ となることから、理論的には、先頭数字が1で始まる頻度は、5から9で始まる頻度の和に等しくなる。同様に、先頭数字が1で始まる頻度は、2と3で始まる頻度の和に等しくなることなどが導かれる。

表3 日本の諸集団の先頭数字頻度分布

数字	日本 ¹⁾	府県 ²⁾	東京都 ³⁾
総数	119	47	54
1	24	23	18
2	0	8	6
3	19	1	6
4	21	1	4
5	14	4	7
6	10	2	6
7	12	2	3
8	8	6	1
9	11	0	3

注1) 1872年から1990年までの日本人口

注2) 1985年国勢調査による都道府県人口

注3) 1985年国勢調査による区市郡支庁人口

さらに、一定の増加率でなくランダムに成長をしている集団であっても、先頭数字の比の最適推定値は、一定の成長率を考える場合の推定値、すなわち、表4の比と同じになると考えられる。

実際、表6は、1から100,000の間の初期人口に対して、ランダムな増加率（-50%から100%

まで）を与え続けたものであるが、6回ランダムな増加率を掛け合わせる頃から、表4の比とはほぼ同じになることがわかる。

したがって、ある数以上の集団があり、その集団が適当な増減を経験すると、あるいはまた、1集団の人口変動を時系列で観察し、先頭数字の頻度分布の観察を続けると、先頭数字頻度漸減の現象が見られるようになると言えるだろう。

表5 部分集団ごとの年増加率一定（0%~10%）の場合における100年後の人口の先頭数字頻度分布

先頭数字	全体	桁数							
		2	3	4	5	6	7	8	
1	292	1	9	60	77	69	68	8	
2	195	1	9	46	48	48	43	0	
3	121	1	9	31	29	27	24	0	
4	95	2	10	17	22	29	15	0	
5	86	2	9	14	25	24	12	0	
6	56	0	9	9	17	13	8	0	
7	58	1	11	14	16	11	5	0	
8	51	0	6	10	19	11	5	0	
9	45	0	12	11	7	14	1	0	

注) 初期条件は、0 - 1,000 の1,000個の一樣乱数

表6 部分集団がランダムな年増加率（-50%から100%）を次々に経験する場合の先頭数字の頻度分布

先頭数字	初期	試行回数									
		1	2	3	4	5	6	7	8	9	10
1	103	193	236	232	260	276	304	282	300	297	310
2	112	132	138	148	152	158	146	164	155	150	154
3	120	134	137	139	131	129	123	125	116	133	122
4	126	127	125	123	120	98	109	107	101	105	102
5	129	93	105	89	89	102	82	84	88	74	77
6	115	104	85	103	79	65	72	60	83	72	71
7	89	84	78	53	59	70	62	70	46	63	66
8	117	69	50	61	59	61	60	59	62	57	50
9	89	64	46	52	51	41	42	49	49	49	48

注) 初期条件は、0 - 100,000 の1,000個の一樣乱数

現実の人口で検討して見よう。表7は都道府県・男女・年齢別の人口を階級幅を50歳とし、5歳間隔で年齢階級をずらしてみたものである。初期人口の先頭数字頻度分布は、男女全体と男女別では理論値と大きく異なる。しかし、0-49歳から50-99歳に進むに連れて、それぞれ表4の理論値に収束していく様子が見てとれる（表4の理論値とのずれを見たカイ2乗値の変化も参照）。すなわち、死亡と移動をより経験すると先頭数字頻度漸減の現象が見られるようになる、と考えられるのである。

なお、先頭数字頻度漸減の法則は、途上国における地域別人口分布の信頼性や年齢別人口分布の正確性のチェックに用いることができるであろう。

表7 都道府県・男女・年齢別人口の先頭数字の頻度分布

(男女計)		(※)											
先頭数字	期待値	年 齢 階 級											
		0 - 99	0 - 49	5 - 54	10 - 59	15 - 64	20 - 69	25 - 74	30 - 79	35 - 84	40 - 89	45 - 94	50 - 99
1	30.1	33.2	36.6	37.2	37.0	35.3	33.6	32.6	30.6	31.3	30.2	30.0	29.8
2	17.6	9.5	4.9	4.9	5.1	5.5	5.5	6.0	8.5	10.0	10.2	12.3	14.0
3	12.5	10.2	7.7	7.4	7.7	7.9	8.9	9.8	10.2	11.3	11.3	12.3	12.8
4	9.7	9.4	8.1	7.9	7.4	8.1	8.7	9.4	10.4	10.0	9.8	10.0	10.6
5	7.9	10.7	11.5	11.7	11.7	12.8	12.8	12.3	11.9	11.7	11.9	10.9	10.0
6	6.7	7.4	7.2	6.8	7.2	6.2	6.6	6.6	6.6	6.2	6.6	7.0	7.7
7	5.8	8.5	9.8	9.8	10.0	10.4	10.2	10.0	8.9	8.5	9.1	8.3	7.2
8	5.1	5.0	6.2	6.6	6.2	6.6	6.6	6.4	6.2	5.7	5.7	4.7	3.8
9	4.6	6.1	8.1	7.7	7.7	7.2	7.0	7.0	6.6	5.3	5.1	4.5	4.0
(カイ2乗値)			20	20	20	20	18	16	10	7	7	4	2
(男 子)													
1	30.1	20.6	13.0	13.0	13.6	14.7	16.6	18.5	21.9	23.8	24.5	26.4	28.3
2	17.6	19.4	19.8	20.0	19.6	20.9	21.3	21.7	20.6	19.8	21.3	19.4	18.9
3	12.5	14.5	17.4	17.2	17.9	18.1	17.9	16.8	16.0	14.9	14.9	13.6	11.5
4	9.7	12.9	13.6	13.6	13.4	13.6	13.2	13.0	12.3	11.5	11.5	12.1	12.1
5	7.9	10.4	13.0	13.4	13.8	13.4	12.6	11.3	10.2	10.0	8.9	8.3	7.9
6	6.7	8.4	9.6	9.4	9.1	8.3	7.7	7.9	7.4	8.3	7.9	7.9	7.2
7	5.8	5.1	6.0	5.3	4.9	4.5	4.3	4.3	3.8	3.6	3.2	3.4	4.3
8	5.1	6.0	5.5	6.2	6.2	5.3	5.3	5.1	5.3	5.3	5.1	5.7	6.4
9	4.6	2.8	2.1	1.9	1.5	1.3	1.3	1.5	2.3	2.8	2.8	3.2	3.4
(カイ2乗値)			19	20	21	19	16	12	7	5	5	3	2
(女 子)													
1	30.1	20.1	12.3	12.1	12.6	13.2	13.8	14.3	17.0	20.6	21.5	25.7	27.9
2	17.6	18.6	18.9	18.1	17.0	17.7	18.7	19.6	20.0	19.6	18.3	18.5	18.3
3	12.5	16.4	16.4	16.4	17.7	17.7	17.9	17.4	17.2	17.0	17.0	16.6	16.4
4	9.7	12.8	14.3	14.3	13.8	13.4	13.6	13.8	13.0	11.5	12.3	11.3	11.3
5	7.9	10.5	13.2	13.6	13.6	14.3	13.6	12.1	11.3	10.4	10.0	8.9	7.9
6	6.7	8.9	11.1	11.3	11.3	11.1	10.0	10.0	9.6	8.5	7.9	7.0	6.8
7	5.8	5.4	6.4	6.4	6.4	6.2	5.7	5.7	5.1	4.7	5.1	4.3	4.5
8	5.1	4.1	5.1	5.3	5.3	4.5	4.9	4.9	4.7	4.7	3.8	3.8	3.2
9	4.6	3.1	2.3	2.6	2.3	2.1	1.7	2.1	2.1	3.0	4.0	3.8	3.8
(カイ2乗値)			21	22	23	22	20	17	13	7	6	3	3

資料) 1985年国勢調査結果より

注) カイ2乗値はそれぞれ割合について計算したものである。