KJHM Vol. 17, No. 1, 2004, 97-108

인구 역 추정에 의한 시조의 연대를 추정하는 수리적 방법

한국항공대학교 명예교수 **유동선**
인하대학교 명예교수 **구자흥**
남서울대학교 교양학부 **이성철**

Abstract

There are so many methods in population estimation such as logistic estimation and compound interest estimation. If we have some pieces of information about population of one tribe, we can estimate progenitor chronology of that tribe used by inverse estimation.

In this paper, we describe several theory of population estimation, and develop mathematical method for estimation progenitor chronology from prior general data and statistical estimation theory. Several examples are illustrated.

0. 서론

통계학에서 대부분의 추정 문제는 장래의 수요, 생산고 등을 추정하는 것으로 되어 있다. 경우에 따라서는 역으로, 현재의 수요 생산고를 통해서 보급시점, 생산 기간 등을 역 추정하는 것이 요구된다. 역 추정은 일반 추정의 역순으로 쉽게 해결되리라 생각되나, 일반 추정보다 더 많은 문제점들이 발생하므로 폭넓은 통계학 지식을 활용하여야· 해결된다. 특히 과거의 데이터가 없는 경우, 현재의 데이터만으로 과거의 사실을 역 추정한다는 것은 그리 쉽지 않다. 여기에는 과거의 개략적인 데이터에 관한 지식을 토대로 통계이론을 이용하여 오차가 적은 추정치를 찾아내야 한다.

1. 인구 추정 법에 관한 고찰

1.1. Logistic 곡선의 유도

인구추정법을 이용하여 현존하는 인구를 근거로 시조(선조)의 연대를 추정하기 위해서는 먼저 오차가 작은, 즉 가장 적합한 추정법을 찾아야 한다. 인구추정법에는 다음과 같은 3가지 방법이 주로 사용된다.

① Logistic 곡선에 의한 인구추정

② 복리법적 인구추정

③ 연속적 인구추정

이들 3가지 인구추정 법에서 우리의 데이터의 형태에 맞는 추정법이 선정되어 가장 오차가 작은 결과를 가져와야 한다.

이제 Logistic 인구추정법에 관하여 살펴보자. 인구가 적을 때 인구는 지수 함수적으로 증가한다. 인구가 증가할수록 억지력이 커져서 인구증가율이 둔해지므로 S자형 경향으로 인구는 증가한다. 이와 같은 인구의 성장 곡선을 Logistic 곡선이라 한다. 이 곡선은 벨기에 수학자 Verhulst 가 인구증가의 법칙을 정식화하는 데서 언급된 것이다. 그는 "인구 증가속도는 $y\left(1-\frac{y}{k}\right)$에 비례한다."고 가정하였다. 여기서 y는 시점 t에서의 현존인구이고, k는 최종적으로 도달된다고 생각되는 인구의 상한값이다.

인구 증가율을 식으로 나타내면 다음과 같은 미분방정식으로 나타내어진다.

$$\frac{dy}{dt}=ay\left(1-\frac{y}{k}\right) \qquad (a\text{는 비례상수}) \tag{1}$$

이 방정식을 y에 관하여 풀기 위해 변수분리하면, 다음과 같다.

$$\frac{kdy}{y(k-y)}=a\,dt,$$

$$\left(\frac{1}{y}+\frac{1}{k-y}\right)dy=a\,dt$$

적분하면 다음을 얻는다.

$$\log\frac{y}{k-y}=at+\log c,$$

$$\frac{y}{k-y} = ce^{at},$$

$$y = (k-y)ce^{at},$$

$$y(1+ce^{at}) = kce^{at}$$

따라서 구하는 y는 다음과 같다.

$$y = \frac{kce^{at}}{1+ce^{at}}$$

$$= \frac{k}{1+me^{-at}} \quad \left(m = \frac{1}{c}\right) \tag{2}$$

이 곡선의 계수 m과 a의 값은 과거의 자료를 토대로 결정한다.

1.2 Logistic 곡선에 의한 인구 추정

인구추정 공식 (2)에 의하여 해방 후의 우리나라 인구추정공식을 구해보자. 식 (1)을 편리하게 사용하기 위하여 dy를 Δy로, dt를 Δt로 바꾸고 $\Delta t = 1$로 하여 선형화하면, 식 (1)은 다음과 같이 된다.

$$\Delta y = a - \frac{a}{k}y \tag{3}$$

여기서 $\frac{\Delta y}{y} = w$ 이고 $-\frac{a}{k} = b$라고 놓으면, 식 (3)은 다음과 같이 선형화 된다.

$$w = a + by \tag{4}$$

상수 a와 b는 최소제곱법으로 구하고 식 (2)의 m값은 주어진 데이터의 t와 y의 몇 개 값의 평균값을 써서 근사적으로 구할 수 있다.

[例 1] 1960년에서 2000년까지 10년 간격으로 주어진 우리나라 인구통계를 써서 인구추정공식 (2)를 결정하기 위하여 표 1을 작성한다.

식 (4)의 계수 a와 b를 구하는 정규방정식은 다음과 같다.

$$\sum w = na + b\sum y,$$

$$\sum wy = a\sum y + b\sum y^2$$

<표 1> 인구자료

연도	t	y_t	Δy_t	$1/y_t$	$w=\Delta y_t/y$	y_t^2	wy_t
1960	0	25.0	7.2	0.040	0.288	625.00	7.20
1970	1	32.2	5.9	0.031	0.183	1036.84	5.89
1980	2	38.1	4.8	0.026	0.126	1451.61	4.80
1990	3	42.9	4.1	0.023	0.094	1840.41	4.03
2000	4	(47.0)	-	-	-	-	-
계		138.2	22.0	0.12	0.691	4953.86	21.92

여기서 다음을 얻는다.

$$0.691 = 4a + 138.2b,$$

$$21.92 = 138.2a + 4954b$$

이 연립방정식을 풀면, $a = 0.5495$, $b = -0.0109$이고 $k = -\frac{a}{b} = 50.413$이다. 여기서 $y = \frac{k}{2} = 25.206$(백만) 이므로, $t=0$ 가까이에 변곡점이 있다. $t = 0, 1, 2$의 평균은 $\bar{t} = 1$이고 $\overline{y_t} = 31.77$이므로 다음을 얻는다.

$$m = \left(\frac{50.413}{31.77} - 1\right)e^{0.5495} = 1.0166$$

따라서 구하는 Logistic 곡선은 다음과 같다.

$$y = \frac{50.413}{1 + 1.0166e^{-0.55t}}$$

그러므로 다음을 얻는다.

$$y_5 = \frac{50.413}{1.065} = 47.336(\text{백만})$$

불행하게도 우리에게는 1960년 이전의 데이터가 없으므로, Logistic 곡선을 써서 1960년 이전의 인구를 추정하는 것은 어렵다. 따라서 Logistic 방법을 써서 과거 우리 민족의 인구 성장을 유추하기란 다소 어려움이 따른다.

2. 우리나라 인구 배가(倍加)기간의 추정

2.1 지속적인 인구증가에서의 배가기간

연간 인구 증가율을 r, 초년도의 인구를 y_0, t년 후의 인구를 y_t라 하면, 인구의 시간에 따른 증가율은 현존하는 인구 y에 비례하므로 다음이 성립한다.

$$\frac{dy}{dt} = ry$$

변수 분리하여 해를 구하면, 다음을 얻는다.

$$y = ce^{rt}$$

$y(0) = y_0$이라 하면, 시점 t에서의 인구 y_t는 다음과 같다.

$$y_t = y_0 e^{rt} \tag{5}$$

이제 증가율 r을 써서 인구가 배로 증가하는 기간을 구한다. 식 (5)에서 배가기간을 t라 하면, 다음이 성립한다.

$$2y_0 = y_0 e^{rt}$$

이것으로부터 다음을 얻는다.

$$2 = e^{rt},$$
$$rt = \log 2$$

따라서 배가 기간 t는 다음과 같다.

$$t = \frac{1}{r} \log 2 \tag{6}$$

y_0과 y_t가 주어질 때 식 (5)으로부터 r을 구하면 다음과 같다.

$$r = \frac{1}{t} \log \frac{y_t}{y_0} \tag{7}$$

이제 1960년과 2000년의 인구수를 써서 배가 기간을 구하여 보자. 식 (7)로부터 다음을 얻는다.

$$r = \frac{1}{40} \log \frac{47.008}{25.012}$$
$$= 0.015774$$

그리고 식 (6)으로부터 다음을 얻을 수 있다.

$$t=\frac{\log 2}{0.0158}$$

$$=43.87$$

즉, 인구의 배가기간은 약 43.9년이다.

같은 방법으로 1960년부터 2000년까지의 우리나라 연중 인구(7월1일 현재)와 배가기간은 다음 표와 같다.

<표 2> 우리나라 인구의 배가기간

연도	인구(천명)	배가 기간
1960	25,012	
1970	32,241	32.9년 38.6년
1980	38,124	43.9년
1990	42,991	66.2년
2000	47,008	75.2년

2.2 개발단계와 인구 증가율

高出産, 多死亡인 저개발국 시절에는 배가기간이 짧으나, 개발도상국 시절에는 小産, 小死亡으로 배가기간이 약간 길어지고, 선진국 시절에는 감퇴인구 경향으로 배가 기간이 아주 길어지고 있다. 인구증가 추세는 경제 개발 및 선진국, 후진국에 따라 달라지고 있다. 인구통계학에서는 이를 3단계로 구분하며, 각 구분마다 배가기간을 구하면 다음 표와 같다.

<표 3> 경제발전단계별 배가기간

단계	저 개발단계	개발단계	선진국단계
인구증가율	높다	낮다	감퇴경향
연도	-1770	1970-1990	1990-2000
연 증가율	0.0254	0.01425	0.0092
배가기간	27.3년	48.7년	75.2년

상기 표에서와 같이 경제와 인구 성장률은 밀접한 관계가 있다. 경제학에서는 사회발전단계를 다음 5단계로 구분하고 있다. 각 단계마다 우리나라 인구증가율의 배가 기간을 표시하면 다음과 같다.

<표 4> 사회발전단계별 배가기간

단계	미개단계	과도단계	도약단계	성숙단계	대중소비단계
기간	-62	62-71	71-91	91-2010	2011-
연 증가율	미지	0.0254	0.014246	0.0006	0.004
배가기간		27.3년	48.7년	115.5년	1732.9년

2.3 우리나라 인구 증가율의 추정

상기 여러 통계표에서 짧은 기간의 인구증가율은 계산하였으나, 인구통계치가 1960년 이전의 자료는 없으므로 몇 백 년, 몇 천년간의 인구증가율을 정확히 추정하는 것은 어려운 점이 있다. 그러나 다행스럽게도 우리가 기억하고 있는 몇 몇 역사적인 연도의 인구와 2000년도의 인구를 써서 인구증가율 및 배가기간을 추정하면 다음과 같다.

<표 5> 인구증가율 및 배가기간추정

역사적 연도	당시인구(천명)	2000년 인구	연인구증가율	배가기간(년)
1919년(기미년)	21,000(남북한)	72,000(남북한)	0.0152	45.6
1945년(해방)	21,000(남한)	47,008(남한)	0.01465	47.3

세계적으로 연인구증가율은 0.0132로 보고, 배가기간을 52.5년으로 보는 것이 통상이다. 상기 표에서 당시 인구는 우리의 귀에 익은 숫자이며, 실제는 1000천명 미만의 오차가 있다고 볼 때 연 인구증가율 0.0132를 부정할 수 없다. 단군이 이 땅을 세운지 4336년이 된다. 연인구증가율을 0.0132로 볼 때 현재의 남북한 인구의 합은 다음과 같다.

$$y = y_0 e^{rt} = 1e^{(0.0132)(4336)} = 71,933,556$$

그러으로 현재 우리들이 알고 있는 남북한 인구 합 7,200만 명에 너무도 근사하다. 이러한 사실에서 세계적으로 통용되고 있는 인구의 연 증가율 0.0132(배가기간 52.5년)는 우리에게도 잘 적용된다고 할 수 있다.

2.4 복리법적 인구증가율

초년도 인구를 y_0, 연평균 증가율을 r, t년 후의 인구를 y_t라 하면 다음이 성립한다.

$$y_t = y_0(1+r)^t \tag{8}$$

$$\frac{y_t}{y_0} = (1+r)^t$$

그러므로 다음을 얻는다.

$$1+r=\left(\frac{y_t}{y_0}\right)^{\frac{1}{t}}$$

즉, 다음을 얻는다.

$$r=\left(\frac{y_t}{y_0}\right)^{\frac{1}{t}}-1 \qquad (9)$$

배가기간을 t라 하면 다음이 성립한다.

$$2y_0=y_0(1+r)^t$$

이것으로부터 다음을 얻는다.

$$(1+r)^t=2$$

그러므로 다음이 성립한다.

$$r=2^{\frac{1}{t}}-1$$

그런데 $t\to\infty$일 때, $r\to 0$이므로 다음을 얻는다.

$$\lim_{t\to\infty}(1+r)^t=\lim_{t\to\infty}[(1+r)^{\frac{1}{r}}]^{rt}=e^{rt} \qquad (10)$$

따라서 식 (8)은 $t\to\infty$일 때 다음과 같이 된다.

$$\lim_{t\to\infty}y_t=y_0\lim_{t\to\infty}(1+r)^t=y_0e^{rt} \qquad (11)$$

그러므로 계속적인 인구증가 시 t년 후의 인구추정공식 식 (5)와 같다. 따라서 시간 t가 충분히 큰 경우는 식 (5)나 식 (8) 중 어느 것을 써서 인구추정을 하여도 큰 차이가 없으나, 실제는 식 (5)가 더 많이 활용되고 있다.

3. 인구의 역 추정에 의한 시종의 연대 추정

상기에서 기술한 연 증가율과 배가기간이 주어지면, 인구추정공식 (5)나 (8)에 의하여 장래의 인구가 추정된다. 역으로 현존하는 인구자료를 토대로 역 추정을 통하여 조상의 연대를 찾는 것도 가능하다. 이제 식(5)와 통계적 해석법을 사용하여 시조(조상)의 연대를 추정

하는 방법을 구하여 보자. 먼저 회귀분석이론을 쉽게 적용하기 위하여 식 (5)를 선형방정식으로 변형한다.

식 (5)의 양변에 로그를 취하면, 다음과 같이 된다.

$$\log y_t = \log y_0 + rt \tag{12}$$

$\log y_t = Y_t$, $\log y_0 = Y_0$이라 놓으면, 식 (12)는 다음과 같이 된다.

$$Y_t = Y_0 + rt \tag{13}$$

이 식은 선형방정식으로 회귀직선의 이론을 쉽게 적용할 수 있다. 인구추정문제에서 정상적인 인구증가법칙에 따르지 않는 부족도 있으므로, 확률분포를 이용하여 신뢰성 검정이 요구되는데 이 경우 회귀분석이론을 적용할 수 있다. 회귀직선 식 (13)에서 t의 함수 Y_t는 고정된 t값에 대하여 정규분포를 따른다. 아울러 충분히 큰 Y_t의 분포 또한 정규분포에 잘 적합되므로 Y_t값에 대한 확률분포를 써서 통계적 분석이 이루어진다.

4. 문제의 제기(실례)

다음에 제시되는 문제는 현재 분쟁을 일으키고 있는 백천 유(劉)씨 시조의 경질에 대한 찬반양론이다. 우리는 이 문제를 통계적인구의 추정법으로 시조의 연대를 밝힘으로서 이 분쟁을 해결하려한다. 백천 유(劉)씨의 현재 인구 수는 2001년 현재 총 26,292명으로 추산된다(남한거주자 수 6,573명, 북한거주자 수 19,719명 추정). 2003년 1월 통계청 발표에 의한 유(劉)씨 본관별 인구는 다음 표 6과 같다.

<표 6> 유씨 본관별 인구(명)

區 分		南 韓	北 韓	合 計
本貫	白川	6,573	19,719	26,292
	江陵	178,913	35,783	214,696
	居昌	19,419	3,883	23,302
	晋州	9,444	1,889	11,333
	其他	28,540	5,708	34,248
全體		242,889	66,982	309,871

* 남한인구는 통계청 발표내용이고, 북한거주 인구는 推算한 것임.

연 인구증가율을 0.0132, 배가기간을 52.5년으로 보면, 시조로부터의 연수와 그 후손의 수는 다음 표 7과 같다.

<표 7> 유씨 시조로부터의 연수와 그 후손의 수

經過年數	780년	960년	1000년	1060년
推定人口	29,614명	318,698명	540,365명	1,045,494명

[例 2] 다음과 같은 두 주장 A, B를 통계적으로 어떻게 분석할 것인가?

주장 A : 현 백천 유씨 회장 측은 백천 유씨의 시조는 고려 건국에 공을 세운 인물 劉相晞로 1,060년 전 인물이라고 주장한다.

주장 B : 유씨 현 족보 및 전체 유씨(강릉파, 거창파, 백천파)의 족보에 의하면, 1,040년 생으로 추정되는 劉荃이라는 1,081년 송나라에서 망명한 분이 유씨의 시조이며, 그의 7대손 國樞(780년 전)가 백천 유씨의 시조이고, 유씨 전체의 시조는 劉荃이다.

A, B 어느 주장도 현 유씨 총인구 30여만, 백천 유씨 그중 26,000여명은 인정되고 있다. 주장 A와 B의 어느 쪽이 옳은 주장인지 통계적 방법으로 입증해보자.

① 주장 B의 결론

유씨 전체의 추정인구는 318,700명이고, 현재 인구는 303,600명으로 추산된다. 인구 역 추정으로 전체 유씨의 시조의 연대는 960년으로 추산되며, 그 중 백천 유씨는 현 인구를 26,300명으로 추산할 때, 백천 유씨의 시조는 인구 역 추정으로 780년으로 추정되므로 주장 A는 부당하다는 것이다.

② 주장 A의 분석

주장 A에 따르면, 시조는 고려 초 인물로 지금부터 약 1,060년 전 인물이 된다. 그렇다면 현존하는 백천 유씨의 후손의 수 y는 다음과 같다.

$$y = e^{(0.0132)(1,060)} = 1,045,500$$

그러므로 100만 명 선에 이르러야 한다. 실제 현재 백천 유씨 인구가 26,300명 선이라는 사실은 후손이 적을 수도 있으니까, 후손의 수로서 시조를 역 추정하는 것은 모순이라고 주장 A측은 들고 나오고 있다.

이런 문제를 인구 통계적 분석에 의하여 해결할 수 있다. (13)식에서 $t=1,060$인 경우, Y_t의 분포를 살펴보자. 우리는 다음 확률을 구해서 해결할 수 있다.

$$P(Y_t \leq \log 26,300)$$

여기서 다음과 같으면 주장 A를 기각하기로 한다(유의수준 5%).

$$P(Y_t \leq \log 26,300) \leq 0.01$$

$t=1,060$(년)에서 최소인구를 0, 최대인구를 평균인구 1,045,500의 2배인 2,091,000명으로 보고 범위를 6.6σ(σ는 표준편차)로 보면 다음을 얻는다.

$$6.6\sigma_y = 2,091,000,$$

$$\sigma_y = 316,818$$

그러므로 다음이 성립한다.

$$Z = \frac{y-\bar{y}}{\sigma_y} = \frac{26,300-1,045,500}{316,818} = -3.217$$

따라서 다음을 얻는다.

$$P(Y \leq 26,300) = P(Z \leq -3.217) = 0.0003 < 0.01$$

이 확률계산에서 $t=1,060$년에서 인구가 26,300명 이하가 될 확률은 0.0003이므로 유의수준 1%에서 A의 주장은 옳다고 볼 수 없다. 그러나 B의 주장은 기각할 근거가 없다.

5. 결론

과거의 데이터가 부족한 상태에서 과거의 사실을 통계적으로 추정하는 문제는 퍽 어려운 일이다. 비록 과거의 데이터가 없다 하더라도 소수의 데이터를 알 수 있는 경우에는 그 데이터 또는 과학적 다른 판단법에 의한 모든 정보를 이용해야 한다.

본 논문에서 제기된 시조의 연대추정은 과거의 데이터가 거의 없고, 최근 데이터만 있는 경우이다. 우리는 인구의 증가법을 도입하여 과거의 일부 데이터를 써서 인구의 성장 법칙을 발견하여 과거를 추정할 수 있었다.

일부 성씨 중에는 정상적이 인구 증가에 따르지 못하고 멸종되거나, 극히 비정상적으로 느린 경우도 있다. 그러나 이런 사실이 있다고 통계이론을 무시하고 주장만 세우는 것은 부당하다. 또한 통계이론에서 충분히 데이터가 커지면 통계적 이론에 잘 적용된다는 사실을

간과해서는 안 된다. 잘못된 주장, 거의 있을 수 없는 일을 단지 확률이 0이 아니라는 사실만을 가지고 그것의 존재를 억지 주장하는 것은 비과학적인 생각이다.

특히 인구의 증가 추세는 어느 정도 인구가 형성되면, 그 후는 인구증가 법칙에 너무나도 잘 따른다. 이것은 통계학의 대수법칙(law of large number)이 입증한다. 그러나 최근 들어 인간이 자연의 법칙을 무시하고 성별감별에 따른 태아의 선택출산, 의도적 출산기피 등 인위적 요소에 의한 변수가 많아 앞으로 인구의 성장이 과거의 인구증가법칙을 따를지는 의문이 아닐 수 없다.

1%도 안 되는 확률이나마 있다 해서 그것의 존재를 주장하는 일은 현대과학에서는 옳다고 판단하고 있지 않음을 이해하고 과학적인 근거를 토대로 이 사회가 화합의 마당으로 성장하였으면 한다.

참고 문헌

1. 구자흥, **인구통계학의 이론과 실제**, 교우사, 2002.
2. 유동선, **통계학 해법 대사전**, 교우사, 2000.
3. 홍종철외 6인, **도표로 보는 통계**, 통계청, 1998.
4. 통계청, **본관별 인구조사통계**, 2003 인터넷 게시판.