족보와 인구조사 자료를 기반으로 한 본관별 혼인과 인구 이동에 대한 연구

DOI: 10.3938/PhiT.24.007

이상훈 · 김범준

Migration of Populations via Marriages in the Past

Sang Hoon LEE and Beom Jun KIM

The study of human migration is fundamentally important and has great potential value. For example, it can augment efficient city planning and protection when faced with pandemic outbreaks. Recently-introduced rich sources of data have truly led to remarkable attempts to characterize modern human (especially short-term) mobility. The scarcity of comparable historical data, however, makes the study of human mobility patterns in the past much more difficult. In this article, we present an analysis of long-term human migration. We use the data record from Korean family books (called “*jokbo*”) to estimate migration patterns via marriages for the past 750 years. We apply two generative models of long-term human migration to quantify the relevance of geographical information to human marriage records in the data, and we find that the different types of geographical distributions of the clans (called “*Bon-Gwan*”) pose rather interesting challenges for the direct application of these models. With the geographical distributions of clans, we quantify the “ergodicity” of clans in terms of how widely and uniformly they have spread across the Korean peninsula, and we cross-compare these results to those obtained using surname data from the Czech Republic. Based on the correlation between ergodicity and mi-

gration in Korea, we identify two different kinds of migration patterns: diffusive and convective. We expect analyses of diffusive versus convective effects in population flows to be widely applicable to the study of mobility and migration patterns across different cultures.

서론

인구 이주 행태에 관한 연구를 물리학적인 혹은 정량적인 관점에서 살펴보는 연구는, 상당히 오래 전인 19세기 초부터 이미 수행되어 왔다.^[1,2] 최근에는 “복잡계(complex system)”와 “빅데이터(big data)”에 대한 연구자들의 관심에 힘입어, 인간 행동에 대한 다양한 실제 자료들이 광범위하게 수집되어 본격적인 연구의 재료가 되고 있다. 매순간의 기록이 실시간으로 대규모로 수집되고 있는 21세기와는 달리, 연구의 대상이 될 자료의 양이 많지 않은 과거의 경우에는 어떤 분석이 가능한 것일까? 다행히 인류는 문자의 발명 이래 끝없이 기록을 남겨왔고, 그 중에는 의도했던 그렇지 않은 인구의 이주 정보가 담겨 있는 것들이 있다.

본 연구팀에서 분석^[3]한 한국의 전통 문화인 족보도 바로 그러한 좋은 예이다. 저자 중 한 명인 김범준 교수와 공동 연구자들은 이전 연구에서 족보에 기록된 며느리들의 본관을 이용하여 과거의 본관 분포를 조사한 바 있다.^[4] 사실, 족보에 기록된 며느리들의 본관은 그 자체로 일종의 지리적인 정보를 담고 있을 수 있고, 이를 이용한다면 본관을 가진 인구의 확률분포함수를 살펴보는 데에서 더 나아가, 과거 인구의 이동 패턴을 유추하는 자료로 사용할 수 있겠다는 깨달음이 본 연구를

저자약력

이상훈 박사는 2010년 KAIST 물리학과에서 연결망에서의 상전이와 임계 현상에 관한 연구로 박사학위를 받았고, 스웨덴 Umeå 대학, 영국 옥스퍼드 대학에서 박사후 연구원을 거쳐 2014년 8월부터 현재 성균관대학교 에너지과학과 BK21+ 연구교수로 재직 중이다.

김범준 교수는 1997년 서울대학교 물리학과에서 초전도 배열의 양자 요동에 관한 연구로 박사학위를 받았고, 스웨덴 Umeå 대학 박사후 연구원, 동 대학 조교수, 아주대학교 조교수와 부교수를 거쳐 2005년부터 현재까지 성균관대학교 물리학과 교수로 재직 중이다.
(beomjun@skku.edu)

REFERENCES

- [1] A. Quetelet, Sur l'homme et le développement de ses facultés: Ou essai de physique sociale (Paris, Bachelier, imprimeur-libraire, quai des Augustins, No. 55, 1835).
- [2] E. Ravenstein, J. Stat. Soc. **46**, 167 (1885).
- [3] S. H. Lee, R. Ffranco, D. M. Abrams, B. J. Kim and M. A. Porter, Phys. Rev. X **4**, 041009 (2014).
- [4] S. K. Baek, P. Minnhagen and B. J. Kim, New J. Phys. **13**, 073036 (2011).

시작하게 된 계기가 되었다. 즉, 족보에 수록된 수많은 결혼 기록이 사실은 결혼으로 인한 인구의 이동을 엿보기에 적합한 좋은 (요즘 유행하는 말로 표현하자면) “빅데이터”인 것이다. 가문 또는 본관을 중요시한 우리 민족의 전통 문화는 현대에 까지 이어져 5년마다 정부 주도로 실시되는 인구주택총조사 중 1985년과 2000년에는 각 행정구역별 본관 분포를 실제로 조사했을 정도이며, 조선시대에도 이미 실제 본관의 유래 지역을 벗어나서 전국적으로 퍼진 본관이 많은 상황에서^[5] 그 자료가 매우 중요하게 본 연구에서 사용되었다.

족보, 인구조사 데이터

본 연구에서 사용된 족보 자료는 이전 연구^[6]에 사용된 10개 문종의 디지털화된 족보이다. 주요 분석에는 기록된 며느리의 수가 가장 많은 첫 번째 족보를 사용하였으며, 나머지 9개 족보에 대한 분석에서도 정성적으로 비슷한 결과를 얻었다. 족보에 기록된 며느리의 출생년도와 본관은 그 무렵에 그 본관을 가진 여성이 해당 족보의 본관 쪽으로 시집을 갔다는 뜻이기 때문에 그 하나하나의 기록을 결혼에 의한 이주 단위로 볼 수 있을 것이다. 예를 들어, 첫 번째 족보에 기록된 학성 이씨(鶴城 李氏) 며느리의 수(출생년도와 관계없이 모두 합한 며느리의 수)는 총 20964명이며 함평 이씨(咸平 李氏) 며느리의 수는 125419명이다. 이것으로부터 함평 이씨 집안으로부터 첫 번째 족보에 기록된 집안으로의 결혼에 의한 인구 이주가 학성 이씨 집안으로부터의 이주에 비해 6배 이상 많았다는 추정 이 가능하다. 그러한 결혼에 의한 인구 이동이 (1) 며느리의 해당 본관의 인구 수와 (2) 시집과 며느리 사이의 거리에 얼마나 영향을 받는지를 정량화하기 위하여 두 가지의 인구 이주 모형을 도입하여 분석하였다.

안타깝게도 (1), (2) 모두 족보를 비롯한 과거의 자료만으로는 추정이 어렵다. 그 이유는 과거의 인구를 정확히 알기 어려우며, 본관이 실제로 위치한 위치가 본관이 유래한 위치와는 많이 다를 수 있기 때문이다. 따라서 각 행정구역별 본관 인구 수 자료를 국가통계포털^[7]에서 1985년, 2000년의 인구주택총조사 자료를 받아서 사용하였다. 인구주택총조사 자료에 등장하는 본관의 인구 수^[1]와 본관의 무게중심점(centroid) 사이의 거리를 이용하면 (1), (2) 모두 추정이 가능하다. 여기서 아주 중요한 가정은 본관의 무게중심점이 실제로 의미가 있을 것이라는 것인데, 사실은 인구주택총조사 자료의 분석 결과 저자

1) 참고문헌 [4]에 의하면 각 본관의 인구 분포의 상대적 비율이 조선 전기부터 일정하다. 따라서 현대의 인구를 이용해 족보에 기록될 당시의 인구를 추정할 것이다.

Fig. 1. Examples of (a) ergodic (Kim from Gimhae) and (b) non-ergodic (Lee from Hakseong) clans. We color the regions of South Korea based on the fraction of the total population composed of members of the clan in the year 2000. We use arrows to indicate the origins of the two clans: Gimhae on the left and Ulsan (“Hakseong” is the old name of the city) on the right. In this map, we use the 2010 administrative boundaries.

중 김범준 교수가 속한 김해 김씨(金海 金氏)는 그림 1(a)와 같이 전국적으로 널리 퍼져 있어서 그 무게중심점이 사실은 큰 의미가 없다는 것을 알 수 있다. 반면에 다른 저자인 이상훈 박사가 속한 학성 이씨(鶴城 李氏)는 그림 1(b)와 같이 울산 지역(“학성”은 울산의 옛 지명이다)에만 주로 분포하기 때문에 무게중심점이 실질적인 의미가 있는 것이다. 이렇게 김해 김씨처럼 전국적으로 퍼진 성씨를 통계물리학적 용어를 따와서 에르고딕(ergodic) 본관 또는 전국구 본관(이하 ‘전국구 본관’), 학성 이씨처럼 특정 지역에만 있는 성씨를 비에르고딕(non-ergodic) 본관 또는 지역구 본관(이하 ‘지역구 본관’)이라고 부르기로 하였다.

연구 방법: 인구 이동 모형과 확산 분석

결혼에 의한 여성들의 인구 이동을 설명할 수 있는 이주 모형으로 대표적인 두 가지를 이용하였다. 인구 이동 또는 이주를 설명하기 위한 모형으로 널리 쓰이는 모형 중 중력 모형(gravity model)^[8]은 꽤 예전부터 쓰인 다음과 같은 공식으로 i 에서 j 로의 인구 이동의 양 G_{ij} 를 설명하려는 시도이다.

REFERENCES

- [5] 노사신, 양성지, 강희맹, 신증동국여지승람(新增東國輿地勝覽), 1530.
- [6] H. A. T. Kiet, S. K. Baek, H. Jeong and B. J. Kim, J. Korean Phys. Soc. **51**, 1812 (2007).
- [7] Korean Statistical Information Service, <http://kosis.kr/>.
- [8] J. Q. Stewart and W. Warntz, J. Regional Sci. **1**, 99 (1958).

Fig. 2. Flux predictions from the special case of the gravity model with $\gamma=0$ and $\alpha=1$, and the radiation models for *jokbo* 1. (a) Scatter plot of the number of clan entries in *jokbo* 1 versus the corresponding centroid in 2000 using the gravity-model flux of women from clan i to clan j and with $\alpha=1$. We compute the line using a linear regression to find the fitting parameter $a_G \approx 6.55(4) \times 10^{-11}$ (with a 95% confidence interval) to satisfy the expression $N_i = a_G G_{ij}$, where G_{ij} is the gravity-model flux and N_i is the total number of entries from clan i in the *jokbo*. (b) We compare the same clan entries using the radiation model. We compute the line using a linear regression to find the fitting parameter $a_R \approx 0.049(2)$ to satisfy the expression $N_i = a_R R_{ij}$, where R_{ij} is the radiation-model flux of women from clan i to clan j and N_i is the total number of entries from clan i in the *jokbo*. In both panels, we color the points using the number of administrative regions that are occupied by the corresponding clans [see Figs. 3(a) and (b)]. The red markers (outliers) in both panels correspond to the clan of *jokbo* 1 (*i.e.*, the case $i=j$).

$$G_{ij} = \frac{m_i^\alpha m_j^\beta}{r_{ij}^\gamma} \quad (1)$$

여기서 m_i 는 i 의 인구수(본 연구에서는 i 라는 본관의 인구 수가 되겠다), r_{ij} 는 i 로부터 j 까지의 거리(본 연구에서는 i 라는 본관의 무게중심점으로부터 j 라는 본관의 무게중심점까지의 거리가 되겠다)이며, α, β, γ 는 각 변수에 지수로 붙는 매개변수이다. 조절 가능한 매개변수가 많다는 점은 모형의 단점일 수 있겠으나, γ 는 커질수록 모형에서 거리의 영향력이 커지게 되는 중요한 의미가 있으며, γ 가 0으로 근접할 경우 거리의 영향력이 없는 상황까지 모사할 수 있다.

최근에 고안된 방사 모형(radiation model)^[9]은 이렇게 매개변수가 많은 중력 모형의 단점을 극복하고자, 단순히 i 와 j 라는 양 끝의 정보만 이용하는 것이 아니라 i 를 중심으로 하여 i 로부터 j 만큼의 거리에 해당되는 원 안에 들어있는 인구의 수 s_{ij} 를 이용하여 다음과 같이 매개변수가 없는 i 에서 j 로의 인구 이동의 양 R_{ij} 을 추정한다.

Fig. 3. Distribution of the number of different administrative regions occupied by clans. (a) Probability distribution of the number of different administrative regions occupied by a Korean clan in the year 2000. (b) Probability distribution of the number of different administrative regions occupied by the clan of a Korean individual selected uniformly at random in the year 2000. The difference between this panel and the previous one arises from the fact that clans with larger populations tend to occupy more administrative regions. [That is, we select a clan uniformly at random in panel (a), but we select an individual uniformly at random in panel (b).] (c) Probability distribution of radii of gyration (in km) for clans in 2000. (d) Probability distribution of radii of gyration (in km) for clans of a Korean individual selected uniformly at random in 2000. The difference between this panel and the previous one arises from the fact that clans with larger populations tend to occupy more administrative regions.

$$R_{ij} = \frac{\Omega_i m_i m_j}{(1 - m_i/N)(m_i + s_{ij})(m_i + m_j + s_{ij})} \quad (2)$$

여기서 Ω_i 는 i 에서 이동하는 총 인구의 수이며, N 은 전체 인구의 수로서 최근 연구에서 유한크기 보정을 위하여 수정^[10]된 분모에 포함된 항이다.

결 과

첫 번째 족보에 대한 분석 결과, 식 (1)의 중력 모형을 신부

REFERENCES

- [9] F. Simini, M. C. González, A. Maritan and A.-L. Barabási, *Nature* **484**, 96 (2012).
- [10] A. P. Masucci, J. Serras, A. Johansson and M. Batty, *Phys. Rev. E* **88**, 022812 (2013).

Fig. 4. (a) Probability distribution of the number of different administrative regions occupied a Czech family name in 2009. Note that the two leftmost bars have heights of 0.17 and 0.03 (with 0.17 at the farthest left), but we have truncated them for visual presentation. (b) Probability distribution of the number of different administrative regions occupied by the clan of a Czech individual selected uniformly at random in 2009. The difference between this panel and the previous one arises from the fact that clans with larger populations tend to occupy more administrative regions. [That is, we select a clan uniformly at random in panel (a), but we select an individual uniformly at random in panel (b).] (c) Probability distribution of radii of gyration (in km) of Czech family names in 2009. (d) Probability distribution of radii of gyration (in km) of Czech family names of a Czech individual selected uniformly at random in 2009. The difference between this panel and the previous one arises from the fact that clans with larger populations tend to occupy more administrative regions. Observe that the distributions in panels (a) and (b) are starkly different from the distributions in panels (c) and (d) from Fig. 3.

(며느리) 측 i 에서 신랑(족보에 해당하는 본관) 측 j 에 적용하여 선형 맞춤(fitting)을 했을 때 $\alpha \approx 1.0749$, $\gamma \approx -0.0349$ 정도의 값(하나의 족보의 경우 신랑 측에 해당하는 m_j 가 상수기 때문에 β 는 의미가 없다)을 얻을 수 있었다. 즉, 그림 2(a)에서 볼 수 있듯이 α 는 거의 1에 가까운 선형 비례, γ 는 거의 0에 가까워서 거리와 무관하다는 결과이다. 이러한 중력 모형의 $\gamma \approx 0$ 라는 결과는 실제로 결혼이 거리와 관계없이 이루어졌다는 해석보다는 널리 퍼진 전국구 본관의 존재로 인해 신랑 측 혹은 신부 측의 위치가 그만큼 불확실하여 결혼이 거리와 관계없이 이루어진 것처럼 보인다는 해석이 타당하다. 실제로 분석에서 사용된 10개의 족보에 해당하는 본관이 모두 전국구 본관이었기 때문에 신랑 측의 위치가 이미 불확실하기 때문에 생긴 것으로 보인다.

Fig. 5. Fraction of ergodic clans and distance scales of clans. For this figure, we use the 3900 clans from the 2000 census data for which we were able to identify the origin locations. (a) Fraction of ergodic clans versus distance to Seoul. The correlation between the variables is positive and statistically significant. (The Pearson correlation coefficient is $r \approx 0.83$, and the p -value is $p \approx 0.0017$.) For the purpose of this calculation, we call a clan “ergodic” if it is present in at least 150 administrative regions. We estimate this fraction separately in each of 11 equally sized bins for the displayed range of distances. The gray regions give 95% confidence intervals. (b) Fraction of ergodic clans versus the distance between the location of clan origin and the present-day centroid. We measure ergodicity as in the left panel, and we estimate the fraction separately for each range of binned distances. (We use the same bins as in the left panel.) The correlation between the variables is positive and significant up to 150 km ($r \approx 0.94$, $p \approx 0.0098$) and is negative and significant for larger distances ($r \approx -0.98$, $p \approx 2.4 \times 10^{-4}$).

또한, 중력 모형과는 달리 거리에 무관한 영역을 매개변수로 조절할 수 없는 식 (2)의 방사 모형을 적용했을 경우에는 그림 2(b)와 같이 잘 맞지 않음을 확인할 수 있다. 그리고 그림 2(a), (b) 공통으로 빨간 점으로 표시한 동성동본혼(同姓同本婚)이 모형 예측이나 실제 자료의 경향성에 비해 훨씬 적게 나왔다는 것으로부터, 1997년 헌법재판소에서 헌법 불합치 판결을 받고 2005년에서야 법적으로 허용^[11]된 동성동본혼을 금기시키는 전통이 족보 기록에 고스란히 드러난다는 것을 알 수 있다.

그렇다면 이렇게 이미 널리 퍼져서 중력 모형이나 방사 모형을 그대로 적용하기 힘들게 만드는 전국구 본관은 어느 정도 있는 것일까? 한 가지 힌트는 그림 1에서 보인 김해 김씨와 학성 이씨의 극단적인 다른 분포이다. 인구주택총조사 자료에 존재하는 모든 본관을 대상으로 그 본관이 한 명이라도 존재하는 대한민국 행정구역 시군구 단위의 수를 세어 본 결과,

REFERENCES

- [11] 동성동본금혼 40년 만에 폐지. <http://www.womennews.co.kr/news/42993> (여성신문).
- [12] J. Novotný and J. A. Cheshire, PLOS ONE 7, e48568 (2012).

비단 그 두 본관에만 한정된 것이 아니라 실제로 한국의 본관들이 대체로 양극단에 몰려 있음을 그림 3을 통해서 알 수 있다. 이것은 그림 4의 체코의 성씨 분포^[12]와 매우 다른 양상이며, 문화, 역사의 차이로 인해 문화권마다 매우 다른 성씨 또는 본관의 분포가 나타날 수 있다는 것을 알 수 있다.

그렇다면 한국의 본관은 어떤 과정을 통해 이렇게 극단적으로 다른 두 가지 형태(전국구, 지역구)를 띠게 된 것일까? 그림 5가 보여주는 분석이 이것에 대한 흥미로운 단서를 제공한다. 각 본관이 동성동본혼을 피하기 위하여 그 본관이 밀집해 있는 곳으로부터 다른 곳으로 자연스럽게 확산(diffusion)된다고 가정하면, 여러 곳에서 발생한 본관이 전국적으로 무작위 확산 과정을 통해 퍼지게 될 것이라는 것을 추론할 수 있다. 그림 5(a)는, 특정 본관의 발생 위치로부터 현재 대한민국의 수도인 서울까지의 거리가 멀수록 그 본관이 전국구일 확률이 높아지는 것을 보여준다. 이것은 실제로 서울이 많은 인구를 흡수하는 역할을 하기 때문에 멀리서 서울로 온 본관일수록 전국적으로 널리 퍼져 있을 가능성이 높다는 자연스러운 확산형 인구 이동의 증거라고 할 수 있다.

하지만, 그림 5(b)를 보면, 본관의 발생 위치로부터 현재의 무게중심점까지의 거리가 멀어질수록 더 전국구일 확률이 낮아지는 구간이 있는데, 이것은 본관이 중간에 아무런 자취도 남기지 않고 먼 거리를 이주했다는 뜻이며 단순한 확산만으로는 한국의 인구 이주를 설명할 수 없음을 의미한다. 이렇게 확산과는 다른 원거리 이주 형태를 대류(convection)형 이주라고 생각할 수 있다. 실제로 조선시대의 수많은 외적의 침입과 일제 강점기, 6.25전쟁과 급속한 산업화 등 격동의 역사를 지닌 한반도의 역사를 생각하면 수궁이 가는 부분이다.

결론

한국의 전통 문화유산인 족보를 단순한 구시대의 유물, 혹은 가부장제의 잔재라고 치부할 수도 있다. 다만, 그렇다고 해서 정말로 독특하고 소중한 문화적 자산인 혼인이나 이주의 기록마저 무시하는 우를 범해서는 안될 것이다. 이 글에서는 족보에 기록된 혼인 기록을 현대의 인구주택총조사 자료와 결합하면 인구 이동과 본관 분포에 관한 흥미있는 연구를 할 수 있음을 소개하였고, 앞으로 더 많은 족보 자료가 디지털화되어 훨씬 더 의미있는 후속 연구를 할 수 있었으면 한다.

아울러 이러한 연구가 조선왕조실록 등의 다른 기록 유산에 대한 통계물리학 내지는 복잡계 데이터 과학적 분석이 좀 더 활발하게 진행되는 것을 촉진하는 촉매제 역할을 할 수 있으면 하는 바람이다. 마지막으로, 모든 문종의 족보 자료가 디지털화되어 분석 대상이 될 수 있었으면 하는 희망을 피력하며, 1985년, 2000년에 이어 올해에도 본관의 지역적 분포가 조사된다면 후속 연구가 크게 기대된다는 말로 글을 맺으려 한다.

그림 저작권 관련 정보

본 글에 사용된 모든 그림은 <http://journals.aps.org/prx/abstract/10.1103/PhysRevX.4.041009> 논문에 사용된 그림이며, 이 논문의 모든 그림은 Creative Commons Attribution 3.0 License (<http://creativecommons.org/licenses/by/3.0/>)에 의해 재사용이 가능합니다.