Family name distributions: Master equation approach

Seung Ki Baek, Hoang Anh Tuan Kiet, and Beom Jun Kim*

Department of Physics, BK21 Physics Research Division, and Institute of Basic Science, Sungkyunkwan University, Suwon 440-746, Republic of Korea

Although cumulative family name distributions in many countries exhibit power-law forms, there also exist counterexamples. The origin of different family name distributions across countries is discussed analytically in the framework of a population dynamics model. Combined with empirical observations made, it is suggested that those differences in distributions are closely related to the rate of appearance of new family names.

PACS number(s): 89.65.Cd, 87.23.Cc

I. INTRODUCTION

Understanding the structure of a population and how it evolves in time has been a critical issue in modern societies. Having started as an economic problem, it soon extended to an environmental one, and states take censuses periodically in order to use the results to design demographic policies. Malthus was the first who presented the mathematical claim, which has been accepted as the fundamental principle of population dynamics, that "population, when unchecked, increases in a geometrical ratio" [1]. In modern terms, he meant the exponential growth tendency of a population of size N with a constant net growth rate r, i.e., $\dot{N}/N \equiv r$ with the time derivative $\dot{N}$; this is often called the Malthusian growth model. Forty years after Malthus published his essay, Verhulst added the idea of the maximal capacity allowed by the environment, K, to the growth model, so that the growth rate r can be negative when the population size N exceeds K; this is referred to as the logistic model [2]. In addition to the inherent variety of dynamics it may exhibit [3], there are also other models reflecting the complexity in population dynamics, one of which has been introduced and termed the θ-logistic model [4–6].

Information on the structure of human population is not only available in many countries, but also very reliable owing to modern census techniques. Various classifications therein allow deeper insight into how subpopulations develop and interact with each other. In this work, we classify people according to their family names (or surnames), and study the family name distribution. These studies can also be important from the viewpoint of genetics in biology, since the inheritance of the family name is often paternal, exactly like the inheritance pattern of the Y chromosome. Furthermore, if one can identify quantitatively the origin of differences in family name distributions across countries, it can provide an understanding of the social mechanism behind the naming behavior in human societies.

The pattern of family name distributions in many countries have already been investigated (see Table I). In Japan, the family name distribution $P(k)$ has been shown to have a power-law dependency on the size k of families, i.e., $P(k) \sim k^{-\gamma}$ with the exponent $\gamma \approx 1.75$ [13]. Later, families in

the United States and Berlin have also been reported to display power-law behavior with the similar exponent $\gamma \approx 2.0$ [11]. The same power-law distributions with exponents $\gamma \approx 1.9$ and $\gamma \approx 1.5$ have also been measured for Taiwanese family names and for names in the Isle of Man, respectively [12]. Extensive research in various countries ranging from Western Europe to South America have again found exponents around 2 (see Refs. [9,10,16] and references therein). In sharp contrast, the Korean family name distribution has been recently investigated, revealing the very interesting behavior of $\gamma \approx 1.0$ [8]. The Korean distribution is very differ-

TABLE I. Summary of the empirical results for family name distributions. The family name distribution function is written as $P(k) \sim k^{-\gamma}$ with k the size of the family, which is the number of individuals who have the same family name. As the sampled population size N is increased, the number N_f of observed family names increases either logarithmically (China and Korea) or algebraically (other countries), giving us two distinct groups.

Region	γ	N_f
China [7]		$\ln N$
Korea [8]	1.0	$\ln N$
Argentina [9]		$N^{0.84}$
Austria [10]		$N^{0.83}$
Berlin [11]	2.0	
France [10]		$N^{0.90}$
Germany [10]		$N^{0.77}$
Isle of Man [12]	1.5	
Italy [10]		$N^{0.75}$
Japan [13]	1.75	$N^{0.65}$
Netherlands [10]		$N^{0.69}$
Norway [22]	2.16	
Sicily [14]	0.46–1.83	$N^{1.0}$
Spain [10]		$N^{0.81}$
Switzerland [10]		$N^{0.73}$
Taiwan [12]	1.9	
United States [11,15]	1.94	
Venezuela [16]		$N^{0.69}$
Vietnam [23]	1.43	$N^{0.27}$

*beomjun@skku.edu

046113-1

ent since the cumulative distribution $P_c(k)$ (the number of family names with more than k members, divided by the total number of family names) becomes logarithmic, which results in an exponential Zipf plot (sizes versus ranks of families) [8]. Throughout the present paper, the rank of a family is defined according to its size in descending order, i.e., the biggest family is assigned rank 1, and the second biggest family has rank 2, and so on. More strikingly, the exponentially decaying Zipf plot suggests that the number of family names N_f found in a population of size N increases logarithmically in Korea (we observe the same behavior for the Chinese family names reported in Ref. [7]), in sharp contrast to the corresponding results for other countries, where N_f grows algebraically with N.

In this work, we investigate the possible mechanism for the differences of family name distributions across countries, by using a simple model of population dynamics. We suggest that the difference originates from the rate of appearance of new family names, which is checked by empirical observations made for the history of Korean family names. In more detail, if new names appear linearly in time irrespective of the total population size, $\gamma=1$ is obtained, whereas if the number of new names generated per unit of time is proportional to the population size, $\gamma \approx 2$ is concluded. We also investigate the family books for several family names in Korea containing genealogical trees, and extract the family name distribution to construct the Zipf plot, revealing that the exponential Zipf plot in Korea has been prevalent for at least 500 years. Family names in other countries such as China, Vietnam, and Norway are newly investigated, and comparisons with existing studies lead us to the conclusion that there are indeed two distinct groups of different family name distributions.

The present paper is organized as follows: We present our master equation formulation, and obtain the formal solution for the distribution function in Sec. II. A detailed analysis is then made in Sec. III for the case of constant name generation rate, and historical observations are also discussed. Section IV is devoted to the other case when branching out from old to new names is allowed, which is followed by a summary in Sec. V.

II. POPULATION DYNAMICS: FORMULATION

We first introduce the master equation in a general form to describe the time evolution of the family size, and then present the formal solution obtained by using the generating function technique.

Let us define the probability $P_{j,k}(s,t)$ for a class (family) to have number $n(t)=k$ at time t given that it started with $n(s)=j$ at time s:

$$P_{j,k}(s,t) = P[n(t)=k|n(s)=j], \qquad (1)$$

which is required to satisfy the initial condition $P_{j,k}(s,s) = \delta_{jk}$ with the Kronecker delta $\delta_{jk} \equiv 1\ (0)$ if $j=k\ (j \neq k)$. The time evolution of $P_{j,k}(s,t)$ is governed by the following master equation:

$$\frac{dP_{j,k}(s,t)}{dt} = \lambda_{k-1}(t)P_{j,k-1}(s,t) + [\mu_{k+1}(t) + \beta_{k+1}(t)]P_{j,k+1}(s,t)$$
$$- [\lambda_k(t) + \mu_k(t) + \beta_k(t)]P_{j,k}(s,t), \qquad (2)$$

where we have made the continuous-time approximation that $P_{j,k}(s,t+1) - P_{j,k}(s,t) \approx dP_{j,k}(s,t)/dt$. For convenience, we take one year as the time unit, and thus the rate variables λ, μ, and β are defined in terms of the annual change of population. The first term in the right-hand side of Eq. (2) describes the process in which the class with $k-1$ members increases its members by 1, which occurs at the birth rate $\lambda_{k-1}(t)$. The second term is for the opposite process that $k+1$ members is decreased to k members, which occurs when one member either dies at the death rate $\mu_{k+1}(t)$ or invents a new family name at the branching rate $\beta_{k+1}(t)$. We consider only the branching process in which a person invents a new name; changing a name from one to an existing one is not allowed in our model. The last term is for the change from k either to $k+1$ or to $k-1$, which occurs when a person is born, dies, and changes name, at rates $\lambda_k(t)$, $\mu_k(t)$, and $\beta_k(t)$, respectively. In this work, we allow birth and the death rates to depend on time, and write them as

$$\lambda_k(t) = k\lambda\phi(t), \quad \mu_k(t) = k\mu\phi(t), \quad \beta_k(t) = k\beta\phi(t). \quad (3)$$

The prefactor k is easily understood since the family with k members has a chance proportional to k to be picked up. We henceforth also assume that $\lambda > \mu + \beta$ to describe a population growing in time.

The solution of the master equation [2] is easily found by using the generating function written as (see, e.g., Ref. [17])

$$\Psi_{j,s}(z,t) \equiv \sum_{k=0}^{\infty} z^k P_{j,k}(s,t), \qquad (4)$$

with the initial condition $\Psi_{j,s}(z,s)=z^j$ [see Eq. (1)]. It is straightforward to get the following partial differential equation for Ψ by combining Eqs. (2) and (4):

$$\frac{1}{\lambda\phi}\frac{\partial\Psi}{\partial t} = (z-1)(z-\bar{\mu})\frac{\partial\Psi}{\partial z},$$

with $\bar{\mu} \equiv (\mu+\beta)/\lambda < 1$, which is written in the simpler form

$$\frac{\partial\Psi}{\partial\tau} = \frac{\partial\Psi}{\partial x},$$

by introducing new variables x and τ as

$$d\tau \equiv \lambda\phi\,dt,$$

$$dx \equiv dz/(z-1)(z-\bar{\mu}).$$

The solution should be written as $\Psi(x,\tau)=g(x+\tau)$, and the functional form of $g(x)$ is determined from the initial condition $\Psi_{j,s}(z,s)=z^j$ (we henceforth set $j=1$, i.e., the class started from only one member):

$$\Psi = 1 + \left[\left(\frac{1}{1-\bar{\mu}} + \frac{1}{z-1} \right) e^{-(1-\bar{\mu})(\tau-\sigma)} - \frac{1}{1-\bar{\mu}} \right]^{-1},$$

where $\tau - \sigma = \int_s^t \lambda \phi(t')dt'$. By expanding the generating function, we reach the desired probability distribution

$$P_{1,k}(s,t) = \begin{cases} 1 - (1-\bar{\mu})(1-\bar{\mu}\eta)^{-1} & \text{for } k = 0, \\ \eta^{k-1}(1-\eta)(1-\bar{\mu}\eta) & \text{for } k > 0, \end{cases}$$

where

$$\eta = \frac{1 - e^{-(1-\bar{\mu})(\tau-\sigma)}}{1 - \bar{\mu}e^{-(1-\bar{\mu})(\tau-\sigma)}} = \frac{1-R}{1-\bar{\mu}R}, \qquad (5)$$

with $R \equiv e^{-(1-\bar{\mu})(\tau-\sigma)}$.

So far we have focused on the size of a class first introduced at a certain time s ($<t$). To derive the overall population distribution observed at time t, one needs to know when each class was introduced. Let $\Pi(s)$ represent the rate at which a class is introduced. If the history begins at $t=0$, the resulting population distribution at time t is given by

$$P(k,t) = \frac{\int_0^t P_{1,k}(s,t)\Pi(s)ds}{\int_0^t \Pi(s)ds}, \qquad (6)$$

where $N_f(t) \equiv \int_0^t \Pi(s)ds$ is the total number of family names at time t.

Although one can think of a further generalization using different time-dependent functions $\phi_\lambda(t)$, $\phi_\mu(t)$, $\phi_\beta(t)$, for the corresponding rate variables in Eq. (3), for simplicity we restrict ourselves only to the identical form $\phi(t)$. Within this limitation, it is noteworthy that our expression for the family name distribution in Eq. (6) applies for a variety of different situations for arbitrary $\Pi(s)$ and $\phi(t)$. For example, the widely used Simon model [11,15,18–20] corresponds to the situation that $\Pi(s) = \text{const}$ and $\phi(t) \propto 1/N(t)$ with the total population $N(t)$. It is to be noted that the use of $\phi(t) \propto 1/N(t)$ introduces an effective competition among individuals: In one unit of time, only some fixed number of individuals are allowed to be born or die, which yields a linear increase of population in time, different from what really happened in human history. Accordingly, we focus below on the case $\phi(t)=1$ to have exponential growth of the population; however, we consider different choices for $\Pi(s)$.

The assumption of time-independent rates with $\phi(t)=1$ [see Eq. (3)] results in $(1-\bar{\mu})(\tau-\sigma)=(\lambda-\mu-\beta)(t-s)$. Without knowing the details of the generation mechanism of new family names, it is plausible to assume that new family names are introduced into the population at the rate

$$\Pi(s) = \alpha + \beta N(s),$$

which contains both the population-independent part (α) and the population-proportional part (βN). The second term $\beta N(s)$ can be easily motivated if we assume that each individual invents a new family name at a given probability β. The population-independent part of the name generation rate

should also be included to describe, e.g., immigration from abroad.

Let us consider a family that started at time s. The expected family size at time t is computed to be [see Eq. (4)]

$$\bar{k}(s,t) = \sum_k kP_{1,k}(s,t) = \left. \frac{\partial \Psi}{\partial z} \right|_{z=1} = e^{(\lambda-\mu-\beta)(t-s)}, \qquad (7)$$

which yields the self-consistent integral equation

$$N(t) = \int_0^t \bar{k}(s,t)\Pi(s)ds,$$

$$= \int_0^t e^{(\lambda-\mu-\beta)(t-s)}[\alpha + \beta N(s)]ds,$$

or, in the differential form,

$$\frac{dN(t)}{dt} = \alpha + (\lambda - \mu)N(t),$$

with the solution

$$N(t) = N(0)e^{(\lambda-\mu)t} + \frac{\alpha(e^{(\lambda-\mu)t} - 1)}{\lambda - \mu} \propto e^{(\lambda-\mu)t}. \qquad (8)$$

As is expected, the population-proportional part βN in $\Pi(s)$ due to the change of names (from the existing one to a new one) has nothing to do with the increase or decrease of the total population, and thus only the population-independent part in $\Pi(s)$ enters $N(t)$.

III. CONSTANT NAME GENERATION RATE

When family names appear uniformly in time and no branching process occurs, i.e.,

$$\Pi(s) = \alpha,$$

we obtain, via change of the integration variable from s to η [see Eqs. (5) and (6)],

$$P(k,t) = \frac{1}{\lambda t}\int_0^1 \eta^{k-1}d\eta = \frac{1}{\lambda tk}\left(\frac{1 - e^{-(\lambda-\mu)t}}{1 - \bar{\mu}e^{-(\lambda-\mu)t}} \right)^k,$$

which yields

$$P(k, t \to \infty) \sim k^{-\gamma} \quad \text{with } \gamma = 1.$$

It is also straightforward to get the number of family names

$$N_f(t) = \int_0^t \Pi(s)ds = \int_0^t \alpha \, ds \sim t, \qquad (9)$$

which, combined with the total population $N(t) \sim e^{(\lambda-\mu)t}$, yields the expression $N_f(t) \sim \ln N(t)$.

Interestingly, the above results are in perfect agreement with what has been found for Korean family name distribution [8]. The cumulative family name distribution becomes logarithmic (i.e., $\gamma=1$), which gives an exponentially decaying Zipf plot where the size is displayed as a function of the rank of the family. Furthermore, one can show directly from
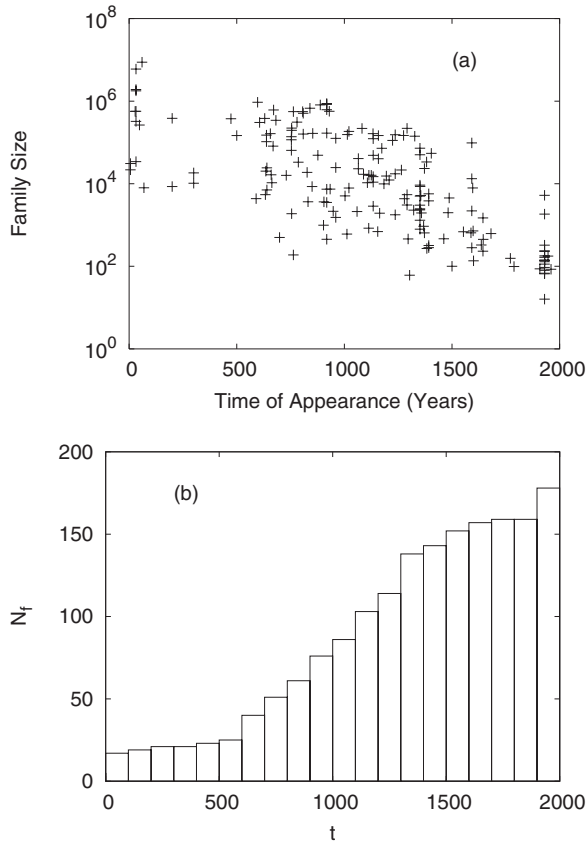
FIG. 1. (a) Each Korean family size versus its time of appearance. The times are collected from the genealogical trees of 178 family names, and the family sizes are from the governmental census data in 1985. The family size grows exponentially since its first appearance as time passes. (b) Number of family names, N_f, as a function of time t in units of years. In Korea, N_f grows approximately linearly in time while the population growth is exponential.

$P(k) \sim k^{-1}$ that the number of family names N_f found for the population size N increases logarithmically, i.e., $N_f \sim \ln N$ [8].

The assumption of the constant rate $\Pi(s)$ of new name generation is very plausible in Korean culture: More or less, it is considered as taboo to invent a new family name, and in Korean history very few names have been introduced; there were only 288 family names in the year 2000 [8]. Furthermore, only 11 names newly appeared between 1985 and 2000, which seems to imply that β in Korea is extremely close to zero. (If we assume that $\alpha = 0$, corresponding to the upper limit estimation of β, we get $\beta \approx 1.8 \times 10^{-8}$ per year.) Korea has preserved its family name system for more than two millennia, and many Korean families still keep their own genealogical trees, from which their origins can be rather well dated. In order to check the validity of the assumption of the constant generation rate of names, we collected information about the origins and sizes of family names from publicly accessible sources. Collecting the sizes and times of appearance for 178 family names, around 60% of those existing, Fig. 1 is obtained. In Fig. 1(a), we show the present size of each family versus the time when it first appeared and it seems to be in accord with the exponential growth in Eq.
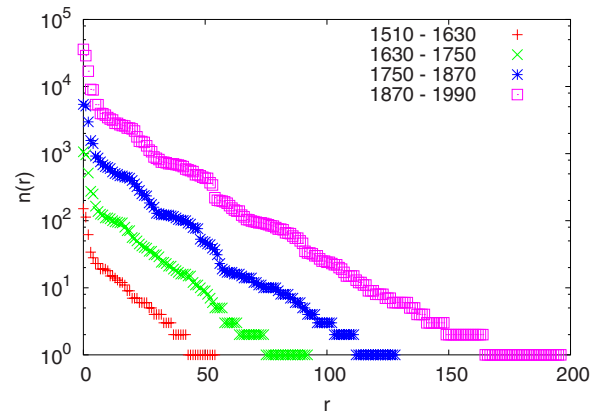

FIG. 2. (Color online) Korean family size $n(r)$ versus the rank r of the family (Zipf plot) extracted from the family names of married women in family books. The exponential decay has been valid for at least 500 years in Korean history.

(7). In Fig. 1(b), we plot the number of family names as a function of time. Although we have included only 178 names, the plot is again in agreement with the linear increase of $N_f(t)$ in Eq. (9) over a broad range of time. We emphasize here that the number of Korean family names increases much more slowly than the total population. We also use several family books containing genealogical trees [21]. Although these books contain only the paternal part of the trees, the family names of women who were married to the members of the family were recorded (in Korea, women do not change their family names after their marriages). We use the information about family names of women at various periods of time to plot Fig. 2. It is clearly seen that the size of the family versus the family rank decays exponentially for a broad range of periods, which confirms that the exponential Zipf plot in Korea has been prevalent for a long time and is not a recent trend. We have shown above that family name distribution of the form $P(k) \sim k^{-1}$ is closely related to the constant generation rate of new names, i.e., $\Pi(s) = \alpha$, which has also been validated from empirical historical observations.

For another example, we present the result of our analysis for Chinese family names in Ref. [7], where 542 262 Chinese are sampled with 1042 family names found. Although only the top 100 Chinese family names are available in Ref. [7], the rank-size distribution (Zipf plot) appears to have preserved an exponential tail for almost a millennium, as shown in Fig. 3(a). Moreover, the number of family names increases logarithmically with the number of people, as depicted in Fig. 3(b), supporting our argument.

We next pursue the answer to the question of how the distribution changes if new names are produced at a rate that is not fixed but grows with the population size.

IV. FAMILY NAME DISTRIBUTION WITH BRANCHING PROCESS ALLOWED

If family names are allowed to branch out, the exponent γ is altered. With β being positive, $\Pi(s)$ is dominated by the
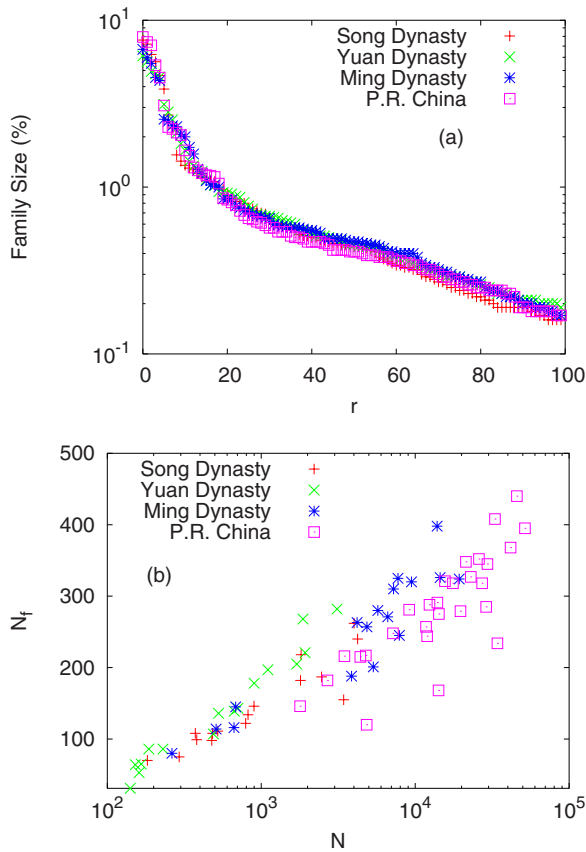
FIG. 3. (Color online) (a) Size distribution of Chinese family names, arranged by their ranks r (the Zipf plot). The exponential shape has been maintained from the time of the Song dynasty (960–1279). (b) Number of people (N) versus the number of family names found therein (N_f), collected in each province of China, showing clearly $N_f \sim \ln N$.

exponential growth in the long run [see Eq. (8)],

$$\Pi(s) = \alpha + \beta N(s) \sim e^{(\lambda-\mu)s},$$

which we use to compute $P(k,t)$ in Eq. (6) as follows:

$$P(k, t \to \infty) \propto \int_0^1 \eta^{k-1} e^{(\lambda-\mu)s} d\eta \propto k^{-\{1+(\lambda-\mu)/(\lambda-\mu-\beta)\}}.$$

Consequently, the family name distribution in the case of $\beta > 0$ has

$$\gamma = 2 + \frac{\beta}{\lambda - \mu - \beta}$$

in agreement with Ref. [12]. It is very reasonable to assume that β is much smaller than $\lambda-\mu$, and we expect $\gamma \approx 2$ in most countries. Indeed, the United States and Berlin have $\gamma \approx 2.0$ [11], which, by using the relation $N_f \sim N^{\gamma-1}$ discussed in Ref. [8], leads to the conclusion that N_f and N are proportional to each other. Of course, one can confirm this linear relation from the direct calculation of the number of family names:
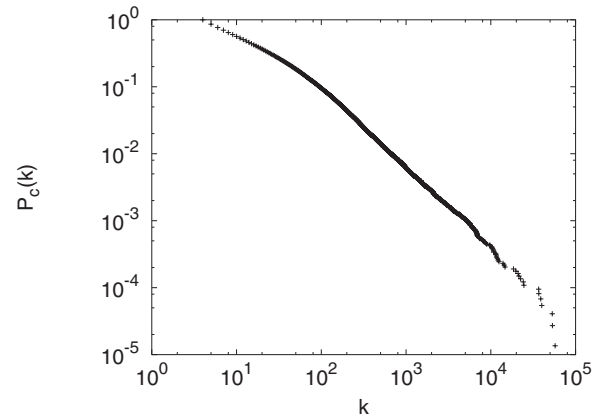

FIG. 4. Cumulative distribution $P_c(k)$ versus the family size k of Norwegian family names, based on a survey in 2007 [22]. The power-law behavior is clearly seen.

$$N_f(t) = N_f(0) + \alpha\left(1 - \frac{\beta}{\lambda-\mu}\right)t + \frac{\beta}{\lambda-\mu}[N(t) - N(0)],$$

which confirms that $N_f/N \to \beta/(\lambda-\mu)$ as $t \to \infty$. The above result of $\gamma = 2 + \beta/(\lambda-\mu-\beta)$ should be used carefully when $\beta \to 0$: If β is strictly zero, one cannot use the assumption $\Pi(s) \sim e^{(\lambda-\mu)s}$, and we recover the result $\gamma = 1$ as previously shown.

From the publicly accessible population information, we estimate that the Swedish population increased at the rate $\lambda - \mu \approx 0.456\%$ per year during 2004–2006. In the same period of time, about 100 new family names were introduced per month, which gives us a rough estimate $\beta \approx 0.015\%$. Accordingly, $\beta/(\lambda-\mu) \approx 0.03$, which makes the assumption we made above very plausible. The number of family names in Sweden is known to be somewhere between 140 000 and 400 000 depending on how we count them. Together with the total population of about 9 million, we confirm that $0.016 \lesssim (N_f/N) \lesssim 0.044$, which is in accord with our expectation that $N_f/N = \beta/(\lambda-\mu) \approx 0.03$.

The empirical findings we have referred to are listed in Table I, from which we suggest two main categories of family name systems: one with $\gamma \approx 1$ and a logarithmic increase of N_f versus N (Korea and China), and the other with $\gamma > 1$ and a power-law increase of N_f (other countries). The latter category has been prevalent in the literature, to which we also add Norway with $\gamma \approx 2.16$ (Fig. 4) and Vietnam with $\gamma \approx 1.43$ (Fig. 5).

We suggest above that the existence of two groups of family name distributions originates from the difference in new name generation rates, which reflects the existence of a very different social dynamics behind the naming behaviors across different cultures. We also point out that, due to the unavoidable simplifications made in our analytic model study, we are not able to clearly explain the spread of γ in the second group of family names. For example, in the history of Vietnam, when a dynasty was ruined many Vietnamese belonging to the fallen dynasty changed their family names into existing ones, which our framework cannot take into account. Another interesting case is the Japanese system.
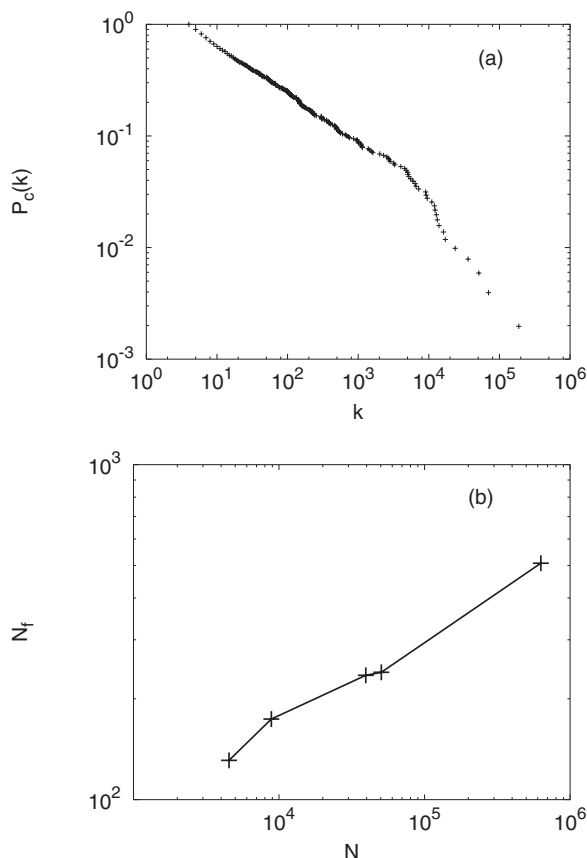
peculiar observation has been found in Sicily: The surname distribution from one of its communes shows $\gamma \approx 0.46$, possibly originating from the effects of isolation [14]; mathematical treatment of this population has not been carried out. Within these limitations of our model study in which various simplifications are made implicitly and explicitly, we strongly believe that such an idealization in general helps one to sensitively check the reality and identify the most important issue from all the ingredients, providing a deeper understanding and insight.

V. SUMMARY

In summary, we analytically investigated the generating mechanism of observed family name distributions. Whereas the traditional approaches from the Simon model are based on implicit assumptions about competition within the population, we instead started from the first principle of population dynamics, the Malthusian growth model. The existence of branching processes in generating new family names was pointed out as the crucial factor in determining the power-law exponent γ: With and without the branching process, $\gamma \approx 2$ and $\gamma = 1$, respectively, were obtained. Genealogical trees collected for Korean family names were analyzed to confirm that the total number of names increased linearly in time, justifying the assumption made in the analytic study. We additionally reported Chinese, Vietnamese, and Norwegian data sets to examine our argument, which, combined with existing studies, lead us to the conclusion that there are two groups of family name distributions on the globe and that these differences can be successfully explained in terms of the differences in new name generation rates.

FIG. 5. (a) Vietnamese family name cumulative distribution, from the phone book of Ho-Chi-Minh City, 2004. (b) The number of family names N_f found in certain numbers of people N. Both show power-law behaviors.

Again, a Japanese family name rarely undergoes the branching process these days, but one finds the algebraic dependency of $N_f \sim N^{0.65}$ [13], which indicates the fact that many Japanese people had to adopt their family names by governmental policy about a century ago. The diversity ensured at the creation appears to be maintained up to now characterizing the Japanese family name system. Consequently, the Japanese name distribution cannot be successfully explained by our model in which the limit $t \to \infty$ is taken. Another

[1] T. R. Malthus, *First Essay on Population* (Macmillan & Co., London, 1926).

[2] J. D. Murray, *Mathematical Biology*, 3rd ed. (Springer, New York, 2003).

[3] S. Strogatz, *Nonlinear Dynamics and Chaos* (Addison-Wesley, Reading, MA, 1994).

[4] B.-E. Sæther, S. Engen, and E. Matthysen, Science **295**, 2070 (2002).

[5] R. M. Sibly, D. Barker, M. C. Denham, J. Hone, and M. Pagel, Science **309**, 607 (2005).

[6] J. D. Reynolds and R. P. Freckleton, Science **309**, 567 (2005).

[7] Y. Yuan, *Chinese Surnames* (East China Normal University Press, Shanghai, 2002) (in Chinese).

[8] B. J. Kim and S. M. Park, Physica A **347**, 683 (2005).

[9] J. E. Dipierri, E. L. Alfaro, C. Scapoli, E. Mamolini, A. Rodriguez-Larralde, and I. Barrai, Am. J. Phys. Anthropol. **128**, 199 (2005).

[10] C. Scapoli, E. Mamolini, A. Carrieri, A. Rodriguez-Larralde, and I. Barrai, Theor Popul. Biol. **71**, 37 (2007).

[11] D. H. Zanette and S. C. Marunbia, Physica A **295**, 1 (2001).

[12] W. J. Reed and B. D. Hughes, Physica A **319**, 579 (2003).

[13] S. Miyazima, Y. Lee, T. Nagamine, and H. Miyajima, Physica A **278**, 282 (2000).

[14] A. Pavesi, P. Pizzetti, E. Siri, E. Lucchetti, and F. Conterio,

Am. J. Phys. Anthropol. **120**, 195 (2003).

[15] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).

[16] A. Rodriguez-Larralde, J. Morales, and I. Barrai, Am. J. Human Biol. **12**, 352 (2000).

[17] E. Parzen, *Stochastic Processes* (Holden-Day, San Francisco, 1962).

[18] H. A. Simon, Biometrika **42**, 425 (1955).

[19] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley,

Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).

[20] K. Yamasaki, K. Matia, S. V. Buldyrev, D. Fu, F. Pammolli, M. Riccaboni, and H. E. Stanley, Phys. Rev. E **74**, 035103(R) (2006).

[21] See, H. A. T. Kiet, S. K. Baek, H. Jeong, and B. J. Kim, J. Korean Phys. Soc. (to be published) for details.

[22] Statistics Norway (http://ssb.no).

[23] Telephone Directory 2004, Ho-Chi-Minh City, Vietnam.