



Comment

First principles theories for last name dynamics

Yosef E. Maruvka^{a,b,*}, Nadav M. Shnerb^c

^a Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA

^b Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

^c Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

Received 4 July 2013; accepted 25 July 2013

Available online 30 July 2013

Communicated by L. Peliti

In many societies, surnames are an indication of a common paternal lineage, a factor of great importance in old and modern civilizations. The ruling class of the Roman republic called themselves patrician, to emphasize their ancestry from the original members of the Senate. Likewise, a major portion of the Bible is devoted to the genealogy of families and tribes.

In modern times, scientists began to develop an interest in the statistics of surnames, i.e., the chance that a single surname, picked at random from a list, is held by m individuals. These statistics may reveal many interesting facts, ranging from mainstream studies, such as assessing the degree of genetic mixing in a population, to esoteric fields like exposing nepotism in the academia [1].

In his review “Surname distribution in population genetics and in statistical physics”, Professor Rossi provides a comprehensive survey of all aspects of the field, with particular emphasis on its relation to the underlying stochastic process. In this process, birth–death events are superimposed on “mutations” (change of surname by individuals) and migration. A review of various technical approaches, like master equations, Fokker–Planck equations, generating function techniques, non-Hermitian quantum mechanics and the renormalization group is also given. In general, this paper provides a detailed review of practically all the important works in the field, from the pioneering studies of Darwin, Galton and Watson to contemporary ideas.

Beyond its practical implications, surname dynamics may be considered the classic example of what is called “neutral” dynamics: a stochastic demographic process in which “mutations” allow for the origination of new taxa, but without selection. As such, the same birth–death–mutation–migration process was presented and studied in many contemporary theories that give more weight to the effect of stochasticity throughout the evolutionary process, and underplay the effects of selection and adaptation. Among these, the neutral theory of molecular evolution [2,3] and the neutral theory of biodiversity [4] have had an immense impact on the fields of population genetics and ecology, correspondingly. Quite recently, a neutral theory of the macro-evolutionary process was also presented and analyzed [5]. A more restrictive model, with birth and mutations but without death, first presented by Yule [6] and by Simon [7], took on a life of its own when it reappeared as the preferential attachment process, a central theme in the theory of scale-free networks [8].

DOI of original article: <http://dx.doi.org/10.1016/j.plrev.2013.06.005>.

* Corresponding author at: Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA.
E-mail address: ymaruvka@jimmy.harvard.edu (Y.E. Maruvka).

Still, a few technical difficulties await a solution. First, the probability distribution function for any stratification of the population is known only for a fixed size population under perfectly neutral dynamics, while for growing populations one can calculate only the average number of families of a given size. Even the variance of these quantities is not known, making a reliable evaluation of the quality of fit quite difficult. Therefore, it is still hard to assess the agreement between empirical data and a theory's predictions. Second, unless one has access to a very large dataset, the characteristic power-law tail of the distribution is very noisy, making the retrieval of the exponent difficult. Indeed, in many cases, one can more successfully fit the typical "shoulder" for small families than fit the tail. The third obstacle relates to the effect of varying growth rates. The theories so far have dealt with a population that grows exponentially at a constant rate (including zero), but the growth rate of realistic populations may vary in time.

These problems, along with others such as the effect of sampling (see the two appendices of the review, and the analysis in [9]) are relevant to all the applications of the birth–death–mutation–migration process, and thus pose a promising research field. Clearly, those engaged in these and other related studies would benefit from an in-depth reading of Professor Rossi's review.

References

- [1] Allesina S. Measuring nepotism through shared last names: the case of Italian academia. *PLoS ONE* 2011;6(8):e21160.
- [2] Ewens WJ. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 1972;3(1):87–112.
- [3] Kimura M. Evolutionary rate at the molecular level. *Nature* 1968;217(5129):624–6.
- [4] Hubbell SP. The unified neutral theory of biodiversity and biogeography. *Monographs in population biology*. Princeton University Press; 2001. p. i–375.
- [5] Maruvka YE, Shnerb NM, Kessler DA, Ricklefs RE. Model for macroevolutionary dynamics. *Proc Natl Acad Sci* 2013.
- [6] Yule GU. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philos Trans R Soc Lond B* 1925;213:21–87.
- [7] Simon HA. On a class of skew distribution functions. *Biometrika* 1955;42(3/4):425–40.
- [8] Barabási A-L, Albert R. Emergence of scaling in random networks. *Science* 1999;286(5439):509–12.
- [9] Maruvka YE, Shnerb NM, Kessler DA. Universal features of surname distribution in a subsample of a growing population. *J Theor Biol* 2010;262(2):245–56.