

Compositional Levy Flights

Selection-like biases emerge in population models with recurrent jackpot events

Authors: Oskar Hallatschek
Genetics 210, 1053 (2018)

*Recommended with a Commentary by Kirill S. Korolev,
Department of Physics and Graduate Program in Bioinformatics,
Boston University, Boston MA*

Random walks find applications in all areas of the natural sciences. First developed to describe Brownian motion and test the foundations of statistical physics [1], random walks are now used to model financial markets [2], organized sports [5], crackling noise [3], tissue morphogenesis [4] and a myriad of other phenomena. Often, one has to consider not only regular diffusion, but also rare events such as stock market crashes that push the system into a different state. While the general theory of such jump-diffusion processes has been developed, there are still many subtleties and open questions in applying this theory to concrete physical phenomena. The featured paper reports remarkable progress in the mathematical description of genetic diversity in situations when the effective number of offspring has a fat tail.

The most widely-used model in population genetics is the Wright-Fisher model. It describes how the relative abundances of genotypes change over time. For simplicity, let us consider just two genotypes: the mutant with relative abundance f and fitness w_m and the ancestral type with relative abundance $1 - f$ and fitness w_a . The model posits that each out of N individuals in the population makes a large number of offspring. Specifically, each mutant organisms makes $w_m W$ offspring and each ancestral type makes $w_a W$ offspring, where $W \gg 1$. The final step consists of randomly sampling N survivors that constitute the next generation. Under these assumptions, the following Fokker-Planck equation describes the temporal evolution of the probability $P(t, f)$ that the mutant has relative abundance f at time t :

$$\frac{\partial}{\partial t} P(t, f) = -s \frac{\partial}{\partial f} [f(1-f)P(t, f)] + \frac{1}{2N} \frac{\partial^2}{\partial f^2} [f(1-f)P(t, f)], \quad (1)$$

where we assumed $N \gg 1$ and $w_m = w_a(1 + s)$ with selective advantage $s \ll 1$. The advection term describes natural selection and the diffusion term describes genetic drift, i.e. changes in f due to the stochasticity of births and deaths. Equation (1) and its extensions

describe the emergence of spread of new mutations and ultimately underline most inferences from genetic data.

The Wright-Fisher model as defined above does not capture the reproduction cycle of most organisms. Yet, numerous studies in population genetics have confirmed that more realistic demographic models reduce to Eq. (1) provided s and N are appropriately defined. This universality is completely analogous to regular random walks whose long-time behavior is completely captured by a single diffusion constant despite potentially complicated microscopic dynamics.

Despite the above-mentioned universality, genetic diversity in several populations was found to be inconsistent with the predictions of the Wright-Fisher model [6, 7]. In such populations, the number (or effective number) of offspring W produced by a single individual has a fat-tailed distribution $P(W) \sim W^{-1-\alpha}$. This unusual situation arises in some marine animals that lay millions of eggs and in viruses that experience periods of exponential growth separated by severe bottlenecks. Effects of exponential growth on the effective number of descendants has been first studied by Luria and Delbruck [8], who found that $P(W) \sim W^{-2}$ and both the first and the second moments of W diverge.

The featured paper by Oskar Hallatschek generalized the Wright-Fisher model to allow each individual to have a random number of offspring with the Luria-Delbruck distribution ($P(W) \sim W^{-2}$). The key result is that the relative abundance of the mutant obeys the following master equation with drift:

$$\frac{\partial}{\partial t} P(t, f) = -\frac{\partial}{\partial f} [s_{\text{eff}} f(1-f) P(t, f)] + \int_0^1 d\tilde{f} [T(\tilde{f} \rightarrow f) P(t, \tilde{f}) - P(t, f) T(f \rightarrow \tilde{f})], \quad (2)$$

$$s_{\text{eff}} = s + \frac{1}{\ln N} \ln \frac{f}{1-f}, \quad (3)$$

$$T(f_1 \rightarrow f_2) = \frac{1}{\ln N} \frac{f_1(1-f_1)}{(f_1-f_2)^2}. \quad (4)$$

Three things are worth noting. First, the diffusion term is replaced by discontinuous jumps. Second, the new terms decrease with population size very slowly as $1/\ln N$ compared to the $1/N$ scaling in Eq. (1). Third, there is non-zero advection that acts as natural selection even when the mutant has no fitness advantage ($s = 0$). Thus, fluctuations are both stronger and acquire non-trivial bias. Hallatschek also obtained the Green function for Eq. (4) in terms of elementary functions, something that is not even possible for Eq. (1).

The origin of the “fictitious” selection lies in the fact that more abundant genotypes are more likely to sample deeper into the tail of $P(W)$ and are, therefore, more likely to benefit from larger “jackpot” reproduction events. As a result, the abundance of rare clones typically decreases. This typical behavior is amended by rare jumps with the opposite bias, which ensures that the relative abundance averaged over independent realizations does not change when mutant has the same fitness as the ancestor ($s = 0$).

Fictitious selection has numerous biological implications. Specifically, it greatly reduces the fixation probability of beneficial mutations and affects inference from time series data on $f(t)$. The technical advances in the featured paper are however likely to be of value beyond population genetics. It is particularly interesting to see how the results generalize for $\alpha \neq 1$. When $\alpha \in (1, 2)$, the variance of $P(W)$ is infinite, and deviations from the Wright-Fisher model, including fictitious selection, are expected. For $\alpha > 2$, however, Eq. (1) should hold.

References

- [1] Einstein, A., *Annalen der physik*, 17, 549 (1905).
- [2] Oksendal, Bernt. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- [3] Colaiori, Francesca, *Advances in Physics* 57, 287 (2008).
- [4] Hannezo, Edouard, et al., *Cell* 171, 242 (2017).
- [5] Gabel, A., & Redner, S., *Journal of Quantitative Analysis in Sports* 8, (2012).
- [6] Li, G., & Hedgecock, D., *Canadian Journal of Fisheries and Aquatic Sciences* 55, 1025 (1998).
- [7] Neher, R. A., & Hallatschek, O. *PNAS* 110, 437 (2013).
- [8] Luria, S. E., & Delbrck, M., *Genetics* 28, 491 (1943).