



## Comment

# Importance of extreme value statistics in biophysical contexts

## Comment on “Redundancy principle and the role of extreme statistics in molecular and cellular biology”

M.V. Tamm <sup>a,b,\*</sup><sup>a</sup> Faculty of Physics, Lomonosov Moscow State University, 119992 Moscow, Russia<sup>b</sup> Department of Applied Mathematics, MIEM, National Research University Higher School of Economics, 101000, Moscow, Russia

Received 9 March 2019; accepted 12 March 2019

Available online 15 March 2019

Communicated by M. Frank-Kamenetskii

Extreme value statistics is an important and beautiful chapter of the general probability theory, see, e.g., [1,2]. In the case of independent identically distributed random variables, which is briefly outlined in chapter 2 of this review [3], the results are strikingly universal. Indeed, for the large number  $N$  of independent variables the distributions of minimum (maximum) converge [1,2] to one of the three stable distributions – Gumbel, Weibull or Frechet, depending on the shape of the lower (upper) tail of the distribution of a single variable. In turn, the typical value of the minimum of  $N$  independent identically distributed random variables  $x_{\min}$  can be estimated [4] by solving the equation  $F(x_{\min}) \approx 1/N$ , where  $F(x)$  is the cumulative distribution of a single variable.

These results have been well established for quite some time, and they are widely known and used both in the pure probability theory and in various applications related to risk management, insurance and financial mathematics. However, this field is much less familiar to people working on the interface between physics and biology. Thus, this review, which is, essentially, an attempt to introduce biophysicists to the topic of extreme value statistics, is extremely welcome. Indeed, there is a variety of biological settings, where the biologically interesting process is triggered at the moment when the first particle (out of many) reaches the target. The examples vary from opening ion channels and triggering avalanches to procreation. The authors argue quite convincingly, that in all those cases it is essential to distinguish between the mean first passage time, which is usually studied in search and narrow escape problems, and the *minimal* first passage time among  $N$  competing independent agents or particles. As a result, the time needed to trigger the process of interest can be significantly shortened: although the reduction in expected time is typically logarithmic in the number of particles, it is still possible to decrease the waiting time by more than an order of magnitude. Indeed, it seems plausible that this reduction in waiting time might be the very reason the nature chooses to produce the redundancy of particles (agents), each of whom is able to trigger the same biological process single-handedly.

It is especially interesting to notice that trajectories of the fastest particles are very dissimilar from typical Brownian trajectories, and bundle around the shortest paths connecting the source of the particles and the target. That is to say,

---

DOI of original article: <https://doi.org/10.1016/j.plrev.2019.01.001>.

\* Correspondence to: Faculty of Physics, Lomonosov Moscow State University, 119992 Moscow, Russia.  
E-mail address: [tamm@polly.phys.msu.ru](mailto:tamm@polly.phys.msu.ru).

if we know just the fact that a given particle is the first to reach the target, we already know a lot about its trajectory. Compare this, e.g., to the fact that statistics of Brownian trajectory returning to the origin might be quite different to the statistics of a generic Brownian trajectory, especially in hyperbolic spaces [5,6], or to the fact that conformations of polymer chains at the moment when they participate in a given chemical reaction are substantially different from generic polymer conformations [7,8]. This dissimilarity between generic and extreme-value trajectories means that the trigger times in various processes discussed in this review should not be sensitive to the minor geometrical details of the search domain (apart, probably, from some simple integral characteristics like the total volume of the domain). On the other hand, they should be sensitive to relatively minor changes of the geometry in the vicinity of the geodesic (shortest path). In particular, putting an obstacle on the geodesic might have a negligible effect on the mean first passage time, but influence the record time significantly. To quantify this influence one needs to construct quantitative theory describing fluctuation properties of the record trajectories, the problem which is stated but not solved in the section 7 of [3]. In relation to this problem I would draw the reader's attention to a recent paper [9], which studies the statistics of random walks over a semi-circle (i.e., in geometry very close to one discussed in section 7 of [3]) and show that in the stretched regime, i.e., when only shortest possible walks are selected the fluctuations of the trajectories are non-Gaussian and belong to the so-called KPZ universality class (see, e.g., [10] for the introduction to this topic).

In conclusion, let us draw the attention of the reader to the fact that almost all examples discussed in the current review correspond to the field of extreme value statistics of *independent* random variables: the agents are moving independently and the first to reach the target triggers the process of interest. The only exception seems to be the study of the endoplasmic reticulum lumen in section 4.3, which contains only brief discussion of qualitative results. Meanwhile, the study of extreme value statistics for *strongly dependent* random variables (see lecture notes [11] for the introduction) is a challenging and rapidly evolving field. Biophysically relevant examples in this case might include fluctuating fibers or membranes. As an example one might think of two membranes described by random functions  $H_{1,2}(x)$  and a process (fusion of two objects) triggered when the membranes first touch. Here values of  $H_{1,2}$  at different  $x$  are random variables, but they are clearly strongly correlated. Another natural example is the statistics of maximal eigenvalue of a random matrix [10,12], the problem which is related to the question of linear stability of many-species ecosystems [13]. Extreme value statistics of strongly dependent variables is a relatively new and rapidly evolving field. One can be sure that in the coming years we will discover more and more examples of its applications in various contexts, including biological ones.

## References

- [1] Kotz S, Nadarajah S. Extreme value distributions: theory and applications. Imperial College Press; 2000.
- [2] de Haan L, Ferreira A. Extreme value theory: an introduction. Springer; 2006.
- [3] Schuss Z, Basnayake K, Holcman D. Phys Life Rev 2019;28:52–79. <https://doi.org/10.1016/j.plrev.2019.01.001> [in this issue].
- [4] Krapivsky PL, Redner S, Ben-Naim E. A kinetic view of statistical physics. Cambridge: Cambridge University Press; 2010.
- [5] Khokhlov AR, Nechaev SK. Phys Lett A 1985;112:156.
- [6] Rubinstein M, Helfand E. J Chem Phys 1985;82:2477.
- [7] Guerin T, Benichou O, Voituriez R. Nat Chem 2012;4:568.
- [8] Guerin T, Benichou O, Voituriez R. J Chem Phys 2013;138:094908.
- [9] Nechaev S, Polovnikov K, Shlosman S, Valov A, Vladimirov A. Phys Rev E 2019;99:012110.
- [10] Kriecherbauer T, Krug J. J Phys A, Math Theor 2010;43:403001.
- [11] Majumdar SN, Pal A. arXiv:1406.6768.
- [12] Majumdar SN, Schehr G. J Stat Mech 2014:P01012.
- [13] May RM. Nature 1972;238:413.