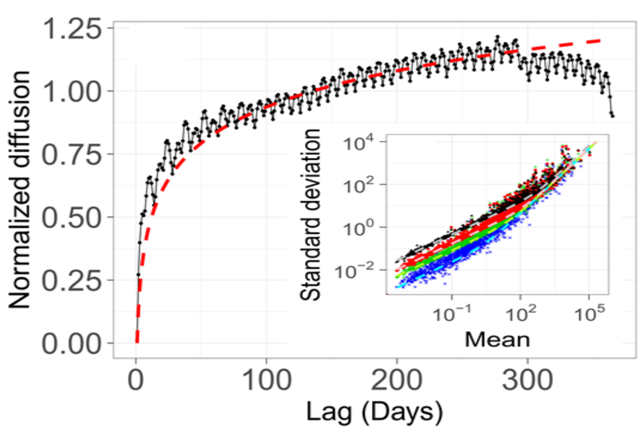
Empirical observations of ultraslow diffusion driven by the fractional dynamics in languages: Fluctuation and dynamical properties of word counts of already popular words

Hayafumi Watanabe
Seijo University, Tokyo, Japan

Ultraslow diffusion (i.e., a kind of logarithmic diffusion; the diffusion where the mean squared displacement is a power of the logarithm) has been extensively studied theoretically, but has hardly been observed empirically. In this research, firstly, we find the ultraslow-like diffusion of the time-series of word counts of already popular words by analysing three different nationwide language databases: (i) newspaper articles (Japanese), (ii) blog articles (Japanese), and (iii) page views of Wikipedia (English, French, Chinese, and Japanese). Secondly, we use theoretical analysis to show that this diffusion is basically explained by the random walk model with the power-law forgetting with the exponent $\beta \approx 0.5$, which is related to the fractional Langevin equation. The exponent β characterises the speed of forgetting and $\beta \approx 0.5$ corresponds to (i) the border (or thresholds) between the stationary and the nonstationary and (ii) the right-in-the-middle dynamics between the IID noise for $\beta = 1$ and the normal random walk for $\beta = 0$. Thirdly, the generative model of the time-series of word counts of already popular words, which is a kind of Poisson process with the Poisson parameter sampled by the above-mentioned random walk model, can almost reproduce not only the empirical mean-squared displacement (MSD) but also the power spectrum density (PSD) and the probability density function. In our presentation, in addition to the above, we will also report on the relationship between the ultraslow diffusion and the fluctuation scaling (FS). The FS is also known as “Taylor’s law” in ecology, is a power-law relation between the system size (e.g., a mean) and the magnitude of fluctuation (e.g., a stan-dard deviation). FS is observed in various complex systems, such as a random work on a complex network, internet traffic, foreign exchange markets, crimes etc. In particular, we will show that the time-scale-independent FS corresponds to essentially a logarithmic diffusion (i.e., a kind of ultraslow diffusion). The figure attached to this abstract shows the logarithmic diffusion in the 1771 main Japanese adjective ensembles (main figure) and the time-scale dependence of the scaling of their fluctuations (inset figure). In the main figure, the black circles are the logarithmic MSD in empirical data and a red dashed curve is theoretical curve (i.e., approximately logarithmic function). Colors and shapes in the inset figure represent the FS of the time series with difference time scales. Black triangles indicate for daily, red circles for weekly, green plusses for monthly and blue crosses for annual time series. Details of the figure are given in the Refs. [1] (the main figure) and [2] (inset figure).



References

[1] Phys. Rev. E. 98, 012308 (2018).
[2] Eur. Phys. J. B. 94,227 (2021).