Zipf's law and criticality in multivariate data without fine-tuning

David J. Schwab, Ilya Nemenman[2], and Pankaj Mehta[3]

***Short Abstract* — Recently it has become possible to measure simultaneous collective states of many biological components, such as neural activities or antibody sequences. Strikingly, the underlying probability distributions of these systems often exhibit a feature known as Zipf's law. They appear poised near a unique critical point, where the extensive parts of the entropy and energy exactly cancel. Here we present analytical arguments and numerical simulations showing that such behavior naturally arises in systems with an unobserved random variable that affects the observed data. The mechanism does not require fine-tuning and may help explain the ubiquity of Zipf's law in disparate systems.**

***Keywords* — Zipf's law, criticality, latent variables, statistical inference**

I. Purpose

In this work [1], we present analytical arguments and numerical simulations showing that Zipf's law [2,3] naturally arises in systems with an unobserved, fluctuating random variable. In particular, we consider a directed graphical model where a hidden variable couples to the observed degrees of freedom. The coupling to the observed data may be heterogeneous and not conditionally independent given the hidden variable. We show analytically that if the distribution of the hidden variable is not too sharply peaked, under general conditions the resulting data will exhibit Zipf's law in the thermodynamic limit. This prediction is tested via numerical simulations of multiple model systems. Finally, we demonstrate this mechanism at work in data using the spiking output of a blowfly motion-sensitive H1 neuron [4] in response to a time-varying input stimulus.

II. Results

A. Analytic results

Zipf-like criticality is a generic property of distributions with fluctuating hidden variables. We show analytically that Zipf's law naturally emerges for large system sizes for probability distributions of the form

$$P(\boldsymbol{\sigma}) = \int d\beta q(\beta) e^{-N\beta\epsilon_N(\boldsymbol{\sigma}) - \log Z(\beta)}$$

where $Z(\beta) = \int d\boldsymbol{\sigma} e^{-N\beta\epsilon_N(\boldsymbol{\sigma})}$ is the partition function and any form of (conditional) interactions among the neurons can be included in the energy. This type of distribution may emerge, for example, when neurons are driven by a common sensory stimulus, or if their physiological state is affected by, e.g. temperature of the preparation.

B. Numerical simulations

In addition to analytical calculations valid in the thermodynamic limit, we numerically test the validity of our analytic result for finite system size in two systems: (a) a collection of non-identical but conditionally-independent spins, and (b) an Ising model with random interactions and fields.

C. Example: H1 neuron

To see our mechanism at work in data, we consider a neural spike train from a single blowfly motion-sensitive neuron H1 [4] stimulated by a time-varying motion stimulus. We discretize time and interpret the spike train as an ordered sequence of spins, corresponding to the absence/presence of a spike in each time bin. Importantly, the probability of a spike in a time bin depends on the stimulus velocity. The rank-ordered plot of spike patterns produced by the neuron is remarkably close to Zipf behavior. We also simulated a refractory Poisson spike train using the same stimulus. The rank-ordered plot for this model still exhibits Zipf's law.

III. Conclusion

Our work demonstrates that Zipf's law can robustly emerge due to the effects of unobserved hidden variables. While our approach is biologically motivated, it is likely to be relevant to other systems where Zipf's law has been observed, and it will be interesting to unearth the dominant mechanisms in particular systems. Our work nonetheless suggests that biological systems operate in a special regime. The system size required to exhibit Zipf's law depends on the sensitivity of the neurons to the hidden variable fluctuations. The system must be well adapted for Zipf-like behavior to emerge at moderate system sizes.

Acknowledgements: This work was funded by NIH grant XX00000. This footnote is optional.

[1]Department of Physics and Lewis-Sigler Institute, Princeton University. E-mail: dschwab@princeton.edu

[2]Departments of Physics and Biology, Emory University. E-mail: ilya.nemenman@emory.edu

[3]Department of Physics, Boston University. E-mail: pankajm@bu.edu

References

[1] Schwab DJ, Nemenman I, Mehta P, http://arxiv.org/abs/1310.0448

[2] M. E. Newman, Contemporary Phys **46**, 323 (2005).

[3] T. Mora and W. Bialek, J Stat Phys 144, 268 (2011).

[4] I. Nemenman, G. Lewen, W. Bialek, and R. de Ruyter van Steveninck, PLoS Comput Biol 4, e1000025 (2008)