POWER LAWS OR NOT POWER LAWS IN ZIPF-LIKE SYSTEMS

Álvaro Corral[1,2]

(1) Centre de Recerca Matemàtica, Edifici C, Campus Bellaterra,
E-08193 Barcelona, Spain.
(2) Departament de Matemàtiques, Universitat Autònoma de Barcelona,
E-08193 Barcelona, Spain.
acorral@crm.es
http://www.crm.es/en/About/People/Researchers/acorral

In Zipf-like systems, individual entities, here called tokens, can be classified into types. Zipf's law is verified when the number of tokens corresponding to any type is a power-law function of the type rank, with exponent close to one. A second formulation consists in performing the statistics of the number of types with the same number of counts, for which Zipf's law states a power-law distribution of type counts, with exponent around two. These two options are equivalent for qualitative purposes, but not in a quantitative sense, due to the discrete nature of the number of counts.

In order to find which of the two formulations (if any) is the correct one, we need to properly fit each of them to empirical data. The difficulties of using maximum likelihood estimation for the rank representation are mentioned, as well as some problems arising from a method proposed recently by Clauset *et al.* to find the onset of a power-law tail. Solving these inconveniences, we find that the rank version of Zipf's law is practically irrelevant for texts, whereas the type-count version fits a much larger set of books from Project Gutenberg.

This has important consequences for the dynamics of the type-token growth (governed by the lowest counts, for which the two versions of Zipf's law are most diverging). We derive the form of the type-token growth curve for a random system fulfilling Zipf's law for the distribution of counts and show that this is not a power law, in contrast with the so-called Herdan's law or Heaps' law. Quite unexpectedly, real books are well fitted by our curve, signaling that, despite the fact words do not happen at random in texts, their first appearance seems indeed to be random.

This presentation is a result of long-term collaborations with, among others, Anna Deluca, Ramon Ferrer-i-Cancho, Fran Font-Clos, and Isabel Moreno.

We acknowledge financial support from FIS2012-31324 and 2014SGR-1307 (Spanish MINECO and AGAUR).

[1] A.C., F. Font, & J. Camacho, *Phys Rev E* 2011.

[2] A.C., A. Deluca, & R. Ferrer-i-Cancho, *ArXiv* 2012.

[3] A. Deluca & A.C., *Acta Geophysica* 2013.

[4] F. Font-Clos & A.C., sub. to *Phys. Rev. Lett.*, *ArXiv* 2014.

[5] A.C. & R. Ferrer-i-Cancho, in preparation, 2015.

[6] I. Moreno, F. Font-Clos, & A.C., in preparation, 2015.