An Extension of the Zipf Distribution

M. Pérez-Casany and A. Casellas

Technical University of Catalonia

The Zipf distribution (Zipf, G. K. 1949) also known as discrete Pareto distribution or Power law distribution is a one-parameter discrete probability distribution obtained by normalizing a p-series. Its support is the set of non-negative integers larger than or equal to one. It is basically characterized by: a) a large probability at one, b) a long right tail and c) a linear behaviour in a log-log scale. It has proved to be useful to fit skewed data from many different areas of research as: linguistics, ecology, insurance, social networking, dynamics of internet, etc. Nevertheless, there are instances where the Zipf model is not flexible enough, basically due to two reasons. On one side, sometimes the empirical and the theoretical probabilities at one differ considerably (usually the empirical is larger). On the other side, the linear behaviour is just observed for values of x large enough, and not from the begining.

In this work a two-parameter extension of the Zipf model is presented. The generalization comes from applying the Marshall-Olkin transformation (Marshall, A. W. and Olkin, I., 1997) to the Zipf model. The inclusion of the additional parameter allows to significatively modify the probabilities at the first positive integer values, while keeping the linear behaviour in the tail of the distribution. In order to show the usefulness of the model presented, several real data sets are analyzed. The fits obtained with the new model are compared with the ones obtained by means of the Zipf distribution, but also with the ones obtained considering other two-parameter models, trucated at zero, such as the Poisson Inverse-Gaussian or the Negative Binomial models.

References

Marshall, A. W. and Olkin, I. (1997) A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, 84, pp 641-652.

Zipf, G. K., (1949) Human behaviour and the principle of least effort, Addison-Wesley Press.