SHANNON ENTROPY AND MIGRATING POPULATIONS

Paulo SARAIVA[1,*]

[1] *Faculty of Economics of the University of Coimbra,*
Centre for Mathematics of the University of Coimbra,
Centre for Business and Economics Research

ABSTRACT

Information theory is the area of mathematics that studies the quantification, storage and communication of digital information. In its genesis is the purpose of quantifying the existing information in random events. This is an area fundamentally established after the work of Harry Nyquist [3] and Ralph Hartley [1] in the 1920s, but mainly by Claude Shannon in the 1940s, in his impressive article *A Mathematical Theory of Communication* [5], situated at the intersection of several disciplines: probability theory, statistics, computer science, statistical mechanics, information engineering and electrical engineering. Applications of fundamental information theory topics include encoding information sources, compressing data (*e.g.*, for ZIP files), encoding channels, as well as error detection and correction. The key concept in information theory is entropy, which is a metric for the degree of causality or uncertainty that random events have. Entropy and amount of information are related as follows: the greater the amount of information, the greater the disorder and the greater the entropy; the smaller the amount of information, the smaller the choice and the smaller the entropy.

This talk starts with a brief introduction to Shannon entropy and its main properties. Recall that

$$\mathrm{H}\left(p_1,\ldots,p_n\right)=\sum_{i=1}^{n}p_i\log_2\left(\frac{1}{p_i}\right)=-\sum_{i=1}^{n}p_i\log_2 p_i.$$

is the so-called Shannon entropy for a random event with n possible states, with probabilities p_i, $i=1,...,n$, where $\sum_{i=1}^{n}p_i=1$.

This measure has been applied in different contexts from what was in its genesis. Indeed, entropy is one of several ways to measure biological diversity, and it is applied by biologists and ecologists under the name of Shannon index (*e.g.*, [2] whcih focus an application relating bird species diversity, plant species and foliage heights).

The main purpose of this talk is to highlight an original application of entropy as a tool to measure the diversity of geographic origins of migrating populations [4]. An example of this application to a restricted set of students at the University of Coimbra is presented. Start by grouping students from a given universe (class, subject, course or university) into categories according to their native place of residence. The diversity measure applied in the present context, being a metric for the diversity of geographical origins of the students, will then result from the computation of

$$\mathrm{H}\left(p_1,\ldots,p_N\right)=-\sum_{i=1}^{N}p_i\log_2\ p_i,$$

where p_i stands for the proportion of students belonging to the i-th category and N is the number of geographic categories considered. This example, if extended to a wider universe, may prove to be an important instrument for the characterization of student migration, with practical consequences.

***K*eywords** Biological diversity · information theory · maximum entropy · population migration · Shannon entropy.

[*] *Corresponding Author's E-mail: psaraiva@fe.uc.pt*

References

[1] Hartley R.V.L., Transmission of information. Bell System Technical Journal, 7(3), 535-563, 1928.

[2] MacArthur R.H. & Macarthur J., On Bird Species Diversity. Ecology, 42(3), 594–598, 1961.

[3] Nyquist H., Certain Factors Affecting Telegraph Speed. Bell System Technical Journal, 3(2), 324-346, 1924.

[4] Saraiva P., On Shannon entropy and its applications, Kuwait Journal of Science 50(3) 194-199, 2023.

[5] Shannon C.E., A Mathematical Theory of Communication. Bell System Technical Journal, 27(3), 379-423, 1948.