Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 132 (2005) 117–129
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Testing the homogeneity of diversity measures: a general framework

M. Salicrú, S. Vives*, J. Ocaña

Department d'estadistica, Universitat de Barcelona, Avda Diagonal 645, 08028 Barcelona, Spain

Received 28 October 2002; received in revised form 12 November 2003
Available online 12 August 2004

Abstract

We discuss the problem of testing the homogeneity of entropy (or diversity) measures, as an extension of asymptotic results concerning the distribution of a reasonable homogeneity statistic. The proposed improvements are based upon corrections similar to those of Satterwhaitte and other authors. As an alternative approach, we also consider a bootstrap procedure. The problem of making multiple comparisons of diversity measures is considered, using either an asymptotic or a bootstrap approach. The use of these methods is illustrated using data, concerning the problem of the extinction of dinosaurs. Finally, we present the results of a simulation study, to determine the relative merits of these methods. The main conclusion is that the corrections to the asymptotic theory and the bootstrap procedure show a good balance between precision and moderate computational cost.

MSC: 62E20; 68U20

Keywords: (h, ϕ)-entropies; Monte Carlo simulation; Asymptotics; Bootstrap

1. Introduction

The thermodynamic concept of entropy was introduced in 1854 by Clausius (1864). In the 1870's, Boltzmann introduced a *statistical* definition of entropy which, he claimed, reduced to the earlier notion of Clausius. Boltzmann was the first in associating a probabilistic meaning to the classical concept of entropy used in Thermodynamics. He can thus be

* Corresponding author.
E-mail address: svives@ub.edu (S. Vives).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.06.018

considered the founder of a branch of Probability and Statistics, known as the Theory of Statistical Information.

In the past century, some authors have associated the probabilistic concept of entropy with the idea of diversity or variability between individuals in a population, in terms of number of classes or categories, M, and the probabilities, $P=(p_1,\ldots,p_M)$, associated with each category: amongst others, Lieberson (1969) in Economy, Lewontin (1972) in Anthropology, and Pielou (1975) in Ecology. This wide use of diversity measures in experimental and observational studies had two main consequences: first, the use of diversity measures on “finite” data posed some questions of statistical inference, concerning the statistical error associated with diversity estimations and second, the variety of application fields and diversity concepts also conducted to a great variety of entropy indexes (Shannon, 1948; Havrda and Charvat, 1967; Renyi, 1961; Arimoto, 1971; Sharma and Mittal, 1975). In other words, there is a need for a *general* inferential theory of entropy indexes.

In that direction, Salicrú et al. (1993) and Pardo et al. (1997) introduced a general entropy functional, with the aim of performing global studies. Under this general model, Shannon, Havrda–Charvat and many other indexes may be considered as (h,ϕ)-entropies, defined as the family of measures of the form

$$H_h^{\phi}(P)=h\left(\sum_{i=1}^{M}\phi(p_i)\right), \tag{1.1}$$

where h and ϕ are real functions twice differentiable with continuity.

Salicrú et al. (1993) studied the asymptotic distribution of (h,ϕ)-entropies under multinomial sampling. If $\hat{P}=(\hat{p}_1,\ldots,\hat{p}_M)$ stands for the maximum likelihood estimator of P, and if $\approx$ stands for “its asymptotic distribution is”, then the following results are obtained:

$$n^{1/2}[H_h^{\phi}(\hat{P})-H_h^{\phi}(P)]\approx \mathrm{N}(0,\sigma_P^2), \tag{1.2}$$

where

$$\sigma_P^2=\left[h'\left(\sum_{i=1}^{M}\phi(p_i)\right)\right]^2\left(\sum_{i=1}^{M}p_i[\phi'(p_i)]^2-\left[\sum_{i=1}^{M}p_i\phi'(p_i)\right]^2\right). \tag{1.3}$$

Given the asymptotic distribution of the (h,ϕ)-entropies, (1.2), it is possible to build general tests and confidence intervals for entropy measures $H_h^{\phi}(P)$ and for the difference of entropies $H_h^{\phi}(P)-H_h^{\phi}(Q)$.

In this context, a test for homogeneity of k groups or populations with respect to a given diversity measure, H_h^{ϕ}:

$$\mathrm{H}_0:\quad H_h^{\phi}(P_1)=H_h^{\phi}(P_2)=\cdots=H_h^{\phi}(P_k),$$

$$\mathrm{H}_1:\quad H_h^{\phi}(P_i)\neq H_h^{\phi}(P_j)\quad i,j=1,\ldots,k \text{ for any } i\neq j$$