1st WORLD CONGRESS
of the Bernoulli Society for Mathematical Statistics and Probability

BERNOULLI
$(\Omega, \mathcal{F}, P)$
SOCIETY ISI

USSR
TASHKENT
1986

ON ENTROPY-BASED DEPENDENCE MEASURES FOR TWO AND THREE DIMENSIONAL CATEGORICAL VARIABLE DISTRIBUTIONS

Virtanen, Ilkka
University of Vaasa
SF - 65100 Vaasa
Finland

Astola, Jaakko
Lappeenranta University
of Technology
SF - 53100 Lappeenranta
Finland

ON ENTROPY-BASED DEPENDENCE MEASURES FOR TWO AND THREE DIMENSIONAL CATEGORICAL VARIABLE DISTRIBUTIONS

VIRTANEN I., University of Vaasa, Vaasa, Finland; ASTOLA J., Lappeenranta University of Technology, Lappeenranta, Finland

Introduction

It is well known that the entropy-based concept of mutual information provides a measure of dependence between two discrete random variables. There are several ways to normalize this measure in order to obtain a coefficient similar e.g. to Pearson's coefficient of contingency, [1], [3], [4], [5], [6].

In our paper we propose and study one way of normalizing the mutual information. There are two factors which make our normalization attractive. First, the coefficient we get possesses a consistent behaviour for a family of test distributions. In a situation where we generate random variables having a "prescribed amount of dependence" among them, we obtain a high degree of compatibility between the entropy-based correlation coefficient and the a priori amount of dependence. Secondly, the definition of the information and the normalization procedure generalize directly to three dimensions. They produce a measure of total dependence among the three variables that possesses the ability to reveal also inverse association or negative dependence between the random variables (even for pure categorical variables).

Two dimensional case

Let X and Y be two discrete random variables with ranges $\{x_1,\dots,x_r\}$ and $\{y_1,\dots,y_c\}$, respectively, having a joint distribution $p_{ij} = P\{X=x_i,\ Y=y_j\}$. The mutual information I_{XY} between X and Y is defined as

(1) $I_{XY} = H_X + H_Y - H_{XY}$,

where H_X and H_Y are the entropies of X and Y

(2) $H_X = -\Sigma_{i=1}^{r} p_{i.} \log p_{i.}$; $H_Y = -\Sigma_{j=1}^{c} p_{.j} \log p_{.j}$,

and H_{XY} is the joint entropy of X and Y

(3) $H_{XY} = -\Sigma_{i=1}^{r} \Sigma_{j=1}^{c} p_{ij} \log p_{ij}$.

The following statements are either direct consequences of the definitions or well known properties of the mutual information:

(4) $0 \le I_{XY} \le \frac{1}{2}(H_X + H_Y)$

(5) $0 \le I_{XY} \le \min\{\log r,\ \log c\}$

(6) $I_{XY} = 0$ iff X and Y are independent

(7) $I_{XY} = \frac{1}{2}(H_X + H_Y)$ iff X and Y are completely dependent.

Now we define the entropy correlation coefficient of X and Y by

(8) $\rho_H = (2I_{XY}/(H_X + H_Y))^{1/2} = (2(1 - H_{XY}/(H_X + H_Y)))^{1/2}$.

The division of I_{XY} by $\frac{1}{2}(H_X + H_Y)$ in (8) is an obvious way to scale the coefficient to [0, 1]. The square root is needed to get a nicely behaving coefficient. The consistent behaviour of the coefficient as a measure of dependence is demonstrated by the following test distributions.

Consider two bi-valued random variables U_1 and V_1 with a joint distribution F_1 : $\{p_{11} = p, p_{12} = 0, p_{21} = 0, p_{22} = 1-p\}$. Then U_1 and V_1 are completely dependent with distributions $\{p_{1.} = p, p_{2.} = 1-p\}$ and $\{p_{.1} = p, p_{.2} = 1-p\}$, respectively. Further, let U_2 and V_2 be independent random variables with the same marginals as U_1 and V_1. The joint distribution of U_2 and V_2 is then $F_2 : \{q_{11} = p^2, q_{12} = p(1-p), q_{21} = p(1-p), q_{22} = (1-p)^2\}$.

Let now X and Y be random variables having a joint distribution $(0 \leq \alpha \leq 1)$:

(9) $$F = \alpha F_1 + (1-\alpha)F_2 = \begin{bmatrix} \alpha p + (1-\alpha)p^2 & (1-\alpha)p(1-p) \\ (1-\alpha)p(1-p) & \alpha(1-p) + (1-\alpha)(1-p)^2 \end{bmatrix} .$$

It would be intuitively natural to argue that the amount of dependence between X and Y is equal to α. Therefore, a proper measure of dependence between X and Y should not be too far from α and it should also be relatively independent of the marginals, i.e. of p.

It is easy to see, that the entropy correlation coefficient has the following properties: ρ_H is scaled to [0, 1] such that 0 indicates full independence and 1 complete dependence between the variables. Further, ρ_H increases almost linearly from 0 to 1 with increasing α, [2, p. 4]. Figure 1 presents a plot of ρ_H as a function of α and p. The plot demonstrates how strikingly well the different requirements set for a dependence measure are satisfied by ρ_H.

Three dimensional case

The information in the three dimensional case, called now total information, is again defined with the help of different order entropies. To get the total information I_{XYZ} between three random variables X, Y and Z, we subtract from the total entropy H_{XYZ} all the lower order entropies:

$$(10)\quad I_{XYZ} = H_{XYZ} - (H_{XY} - H_X - H_Y) - (H_{XZ} - H_X - H_Z) - (H_{YZ} - H_Y - H_Z) - H_X - H_Y - H_Z$$
$$= H_{XYZ} - H_{XY} - H_{XZ} - H_{YZ} + H_X + H_Y + H_Z .$$

The total information I_{XYZ} satisfies the following properties, [2, p. 5-6]:

(11) $-\frac{1}{3}(H_X + H_Y + H_Z) \leq I_{XYZ} \leq \frac{1}{3}(H_X + H_Y + H_Z)$

(12) $I_{XYZ} = 0$, if X, Y and Z are mutually independent

(13) $I_{XYZ} = \frac{1}{3}(H_X + H_Y + H_Z)$, iff X, Y and Z are completely (positively) dependent

(14) $I_{XYZ} = -\frac{1}{3}(H_X + H_Y + H_Z)$, iff X, Y and Z are completely negatively (or inversely) dependent.

By complete positive dependence we mean that for each i, j and k there is at most one pair (j,k), (k,i) and (i,j), respectively, such that $p_{ijk} > 0$. Complete inverse dependence is said to exist if for each pair (i,j), (j,k) and (k,i) there is exactly one k, i and j, respectively, such that $p_{ijk} = 1/m^2$, where $m = \min\{\#(p_{i..} > 0), \#(p_{.j.} > 0), \#(p_{..k} > 0)\}$, [2, p. 5-8.].

The entropy correlation coefficient for a three-dimensional distribution is defined as

(15) $\rho_H = (3I_{XYZ} / (H_X + H_Y + H_Z))^{1/3}$.

The behaviour of ρ_H in three dimensions can be analyzed in an analogous way as in the two-dimensional case. We construct three distributions with equivalent marginals {p, 1-p}, the first exhibiting complete independence, the second complete positive dependence and the third complete inverse dependence. The two test distributions $F = \alpha F_1 + (1-\alpha)F_2$ and $G = \beta F_1 + (1-\beta)F_3$ now possess a prescribed amount of dependence, viz. α (positive dependence) and $-\beta$ (negative dependence), respectively. The analysis shows, [2, p. 9], that the behaviour of ρ_H as a function of α (or β) is quite consistent except in those special cases where α is small and the marginals are highly asymmetric.

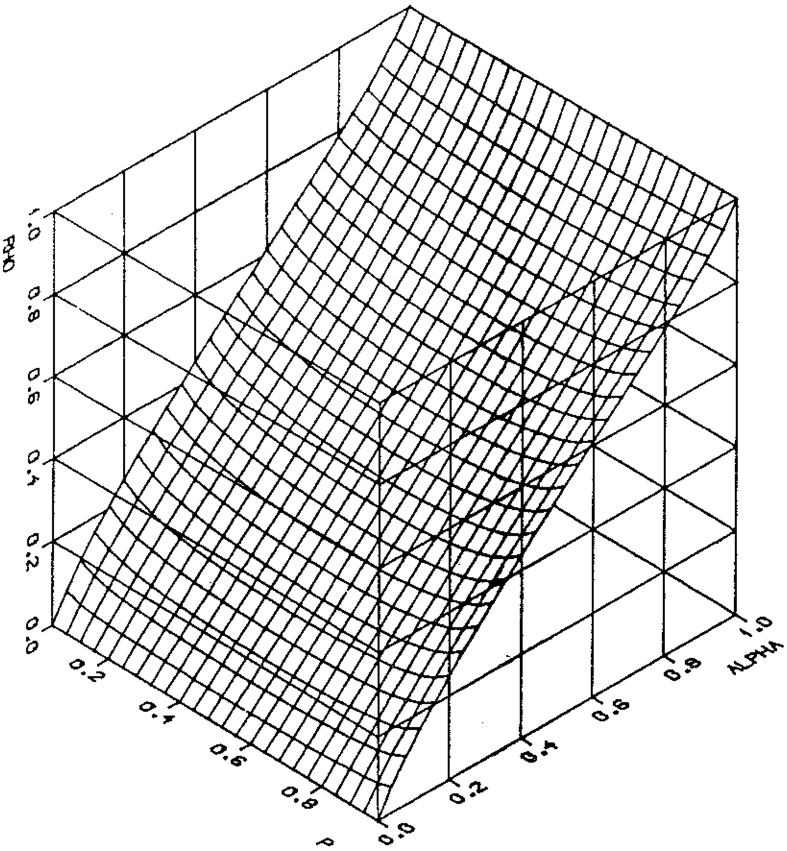

Figure 1. Entropy correlation coefficient ρ_H as a function of α and p.

R e f e r e n c e s

1. Aczél J., Daróczy Z. On measures of information and their characterizations. New York: Academic Press, 1975.
2. Astola J., Virtanen I. On the use of entropy in measuring dependence in two and three dimensions. Manuscript, submitted to Communications in Statistics, 1986, 14 p.
3. Kullbac S. Information theory and statistics. New York: John Wiley & Sons, 1959.
4. Preuss L.G. A class of statistics based on the information concept. Communications in Statistics, Theory and Methods, 1980, Vol. A 9, No 15, p. 1563-1586.
5. Shannon C.E. A mathematical theory of communication. Bell System Technical Journal, 1948, Vol. 27, p. 379-423.
6. Theil H. On the use of information theory concepts in the analysis of financial statements. Management Science, 1969, Vol. 15, No 9, p. 459-480.