Scale–Invariant Correlation Theory

Wassily Hoeffding

Introduction.

The problem of correlation may be described as the investigation of those properties of multivariate distributions which characterize these distributions, *i.e.*, do not occur for univariate distributions. These properties depend above all on the relationships of the variables to each other. From the totality of those properties which belong to the topic of correlation one particular class of properties will be considered more closely.

In dealing with statistical material in which there are *quantitative* characters (random variables) $\xi, \eta, \ldots$, it may be useful *to change the original scale of the variables*. Such a change of scale may be necessary regardless of whether the characters are determined by measurement or by counting. In dealing with measured characters, one might possibly change over to a different unit of measurement (say, from millimeters to centimeters). For any quantitative characters – determined by counting or measurement – it can be advantageous to convert one or more of the variables $\xi, \eta, \ldots$ from the original scale to, say, logarithms. Generally such a *change of scale* is expressed by a transformation

$$\bar{\xi} = f(\xi), \quad \bar{\eta} = g(\eta), \ldots \tag{M}$$

where we shall assume that the transformation (M)

1. is one-to-one (*i.e.*, single-valued and uniquely invertible),
2. leaves the ordering of the variables unchanged, so that *e.g.* $\xi_1 < \xi_2, \bar{\xi}_1 = f(\xi_1), \bar{\xi}_2 = f(\xi_2)$ always imply $\bar{\xi}_1 < \bar{\xi}_2$.

If the variables $\xi, \eta, \ldots$ are continuous and if $f(\xi)$ and $g(\eta)$ are continuous and piecewise differentiable – as we shall assume in what follows – then these two conditions are equivalent to the derivatives $f'(\xi)$, $g'(\eta), \ldots$ being finite, not equal to zero, and positive[1].

[1]See H. Bruns, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*. Berlin 1906, p. 126.

N. I. Fisher et al. (eds.), *The Collected Works of Wassily Hoeffding*

In what follows if we speak of *changes of scale* as such, we shall mean transformations of the form (M) which satisfy Assumptions 1 and 2.

Now if we convert from the original variables $\xi, \eta, \ldots$ to the new variables $\bar{\xi}, \bar{\eta}, \ldots$ by a change of scale, then the distribution of the new variables will be different. Some of the properties of the original distribution will be preserved in the new variables, while others will be affected by the change of scale. *Correspondingly all the properties of a multivariate distribution which pertain to the topic of correlation are divided into two classes, depending on whether or not they are invariant to arbitrary changes of scale.*

Thus, for example, the path of a regression curve is dependent on the particular scale. Also, the related property of not being correlated[2] is bound up with the scale. For example, in the following probability distribution there is no correlation of ξ with η nor of η with ξ. But if we replace ξ by $\bar{\xi}$, putting the value $\bar{\xi} = 2$ for $\xi = 1$, while otherwise keeping $\bar{\xi} = \xi$, then the conditional expected value of $\bar{\xi}$ is no longer the same for all η.

η \ ξ	-1	0	1
-1	0	$\frac{1}{5}$	0
0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
1	0	$\frac{1}{5}$	0

In contrast, *stochastic independence* is a *scale invariant* property. Thus if the characters ξ and η are independent of each other, then the probability of the combination of characters (ξ, η) is equal to the product of the probabilities of the individual variables ξ and η taken by themselves, and for a transformation (M) which applies only to the individual characters, the product representation is preserved (as will be explained in more detail in §1 below).

Just as for stochastic independence the "*degree of stochastic dependence*" of two variables will be described as a scale invariant property. Then the actualization of this concept must correspond so that the variables $\bar{\xi}, \bar{\eta}$ introduced by the change of scale (M) are neither more "strongly" nor more "weakly" stochastically associated than the original variables ξ, η. (In contrast, we will generally assign an entirely different "degree of dependence" to the variables introduced by a transformation of the form $\bar{\xi} = f(\xi, \eta), \bar{\eta} = g(\xi, \eta)$ than to the original variables.) Correspondingly we will demand of a quantity intended to serve as a measure of stochastic dependence of variables that it be *invariant to changes of scale.*

First, one would place such a condition on a genuine index of dependence,

[2] In the sense of A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Berlin 1925, p. 33, "η is not correlated with ξ" means geometrically that the regression line of η on ξ is a line parallel to the ξ-axis.